

# **Remote sensing of environmental variables for mapping malaria distribution: the case of Vhembe District Municipality, South Africa**

Oupa Ermos Malahlela

Submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy  
(Geo-Informatics)

In the Faculty of Natural and Agricultural Sciences  
University of Pretoria

Pretoria

June 2019



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA



**Remote sensing of environmental variables for mapping malaria distribution:  
the case of Vhembe District Municipality, South Africa**

Student Name: Mr. Oupa Ermos Malahlela

---

Student Number: 10064568

Supervisors: Dr. Jane Mukarugwiza Olwoch

---

Dr. Clement Adjorlolo

---

Department: Geography, Geoinformatics and  
Meteorology

Qualification: PhD Geo-Informatics



## Declaration

I, **Oupa Eremos Malahlela**, declare that the thesis, which I hereby submit for the degree of ***PhD Geo-Informatics*** at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: \_\_\_\_\_

Date: \_\_\_\_\_



## **Disclaimer**

This thesis mainly adopts the publication style of writing for each analysis chapter. The thesis comprises of seven (7) chapters, with four (4) objectives that are presented in a paper format for publication. Consequently, three (3) articles have been published in peer-reviewed journals, covering objectives 1, 3 and 4. One article on objective 2 and literature review chapter are still under preparation for submission.

All work done in this study is the original contribution and efforts initiated by the student as the principal investigators.



## Acknowledgments

I would like to express my earnest gratitude to my wife, Nokubonga, who has been very supportive throughout the duration of the study.

I thank my parents Phetola and Koranta for their love and care they showed me from birth until today. I appreciate the person that these two beloved parents made me to be today.

I appreciate the academic foundation laid by the good old teachers, Mr. Joe Moila and Mr. Teenage Malegopo who have since become my pillar of strength and motivation.

To Dr. Paidamwoyo Mhangara, a supportive manager of Research Applications and Development Group, at the South African National Space Agency. I wish to acknowledge his contribution towards supporting this PhD work, both in terms of funding motivation, study arrangements and his for his continued understanding.

I extend my gratitude to Mr. Morwapula Mashalane and Mr. Mahlatse Kganyago for support during the fieldwork design for this PhD. I would also like to thank all Earth Observation team members at SANSa for their continued support in terms of data, time and IT infrastructure.

I wish to thank Dr. Moses Azong Cho and Dr. Abel Ramoelo for laying solid foundation in terms of science of remote sensing and its applications.

The best for last: To my two best supervisors, Dr. Olwoch and Dr. Clement Adjorlolo: This work would not have been accomplished without your supervision, your patience, guidance and contribution to each chapter presented in this manuscript.



## Publications and Manuscripts

The following list of papers and manuscripts were derived from this thesis and include the first author (myself), second and third authors (supervisors) and other co-authors.

Malahlela O.E., Adjorlolo C., Olwoch J.M. (**in preparation**). Applications of remote sensing for malaria mapping: challenges and opportunities. This manuscript relates to the literature review done for this research.

Malahlela O.E., Adjorlolo C., Olwoch J.M. (**2019**). Integrating geostatistics and remote sensing for mapping the spatial distribution of cattle hoofprints in relation to malaria vector control. *International Journal of Remote Sensing*. This paper relates to objectives 1 of the thesis.

Malahlela O.E., Adjorlolo C., Olwoch J.M. (**2019**). Comparison of the Sentinel-2 multispectral and hyperspectral vegetation indices for mapping LAI in malaria-prone heterogeneous semi-arid environment of Southern Africa. *GIScience and Remote Sensing*. This paper relates to objective 2 of the thesis.

Malahlela O.E., Olwoch J.M., Adjorlolo C. (**2018**). Evaluating Efficacy of Landsat-Derived Environmental Covariates for Predicting Malaria Distribution in Rural Villages of Vhembe District, South Africa. *EcoHealth*, 15(1): 23 – 40. This paper relates to objective 3 of the thesis.

Malahlela O.E., Adjorlolo C., Olwoch J.M. (**2019**). Mapping the spatial distribution of *Lippia javanica* (Burm. f.) Spreng using Sentinel-2 and SRTM-derived topographic data in malaria endemic environment. *Ecological Modelling*, 392: 147 – 158. This paper relates to objective 4 of thesis.

## Abstract

Malaria is one of the deadliest parasitic diseases common in the warm, moist tropics of the world. *Plasmodium falciparum* is the causative agent of malaria in human, and it is transmitted to human through the bite of a female *Anopheline* mosquito especially in Africa. It is estimated that approximately 490 000 deaths from malaria were reported in 2015, with over 90% of such cases reported in Sub-Saharan Africa. The global elimination of this disease is one of the most expensive investments that demands cooperation of various disciplines and entities including health, science, technology, political leadership, finance and general country's administration. In recent years, attempts were made to characterize the potential breeding sites of *Anopheline* mosquitoes in order to set baseline for elimination strategies. On the other hand, there is an urgent need to refine malaria mapping methods for updated early-warning and target systems. Mapping of malaria in relation to various environmental and topographical factors is highly advantageous and complements the hospital malaria count and clinical approaches because such environmental-based methods take into account changes in malaria transmission as a result habitat and seasonal and long-term changes in climate and environment. In addition, in very remote areas where there is limited or no access of malaria count data the spatial mapping of malaria becomes crucial for understanding malaria distribution in such isolated areas. The traditional mapping of potential malaria vector habitats is done through field surveys, which are often laborious and time-consuming. The advent of high resolution datasets from earth observation satellites, offers opportunities for accurate mapping of malaria vector habitats for wide areas. The aim of this study was to map the potential habitats of malaria vector (*An. arabiensis* and *An. fenestus*) using remote sensing approaches in Vhembe District Municipality of South Africa. Both Landsat TM and Sentinel-2 datasets were used in the study area. Findings from this study indicate that the distribution of *P. falciparum* is positively correlated to vegetation moisture and greenness. On the other hand, remote sensing data such as that of Sentinel-2 has shown high correlation to the distribution of cattle hoofprints which are known habitats of malaria vectors. It has also been big data shown in this study that the remote sensing spectral indices such as those based on broadband reflectances are paramount to characterizing the resting places of *An. complex*. In an effort to contributing to the indigenous knowledge system (IKS) for repelling malaria vectors common at the study area, this study also showed the feasibility of high spatial resolution Sentinel-2 to map the distribution of *Lippia javanica* used commonly in the Vhembe District. Findings from this study will give insight into the potential habitats of malaria-causing mosquitoes and aid in efforts aimed at eliminating malaria in the Vhembe District.



## Kakaretšo

Bolwetši bja malaria ke bjo bongwe bja malwetši a kotsi kudu ao a bakwago ke kokwanahloko dinageng tše go fišago le mo go le mo go nago le dipula gona. Kokwanahloko ya *Plasmodium falciparum* ke yona yeo e bakago malaria, elego bolwetši bjo bo fetelago mothong ge a longwa ke monang wa sesadi wa lešika la *Anopheles* kudu-kudu Afrika. Go begilwe mahu ao e kabago batho ba 490 000 ao a bakwago ke malaria ka 2015, mo go ona 90 lekgolong ya mahu ao go begwa gore a bile dinageng tša Borwa-bja-Afrika. Maiteko a go fediša leuba la malaria ke le lengwe la dipeeletšo tša go bitša kudu lefaseng, gomme go hloka tirišano ya mekgatlo e fapanego go akaretša tša maphelo, thutamahlale, thekinolotši, boetapele bja tša dipolotiki, tša ditšhelete le tirišo-taolo ya selegae. Mengwageng ya morago bjale, go bile le maiteko a go swaya mafelo ao go ona menang ya *Anopheles* e dulago le go hwetšagalago gona bakeng sa go hloma melao bakeng sa phedišo ya malaria. Ka lehlakoreng le lengwe, ke mo go akgofilego go kaonefatša mekgwa ya go thala mebapa ya malaria bakeng sa tshepedišo ya temošo ya pele. Mokgwa wa kgale wa go thala mebapa ya mafelo ao menang yeo e bakago malaria e dulago gona ke ka go dira dinyakišišo ka go sepela ka maoto, e lego mokgwa wo o lapišago le gore ja nako. Go ba gona ga disatalaete tše di kaonefaditšwedigo go bula menyetla ya go thala mebapa ya menang yeo e bakago menang ka tsela ye nepagetšego kudu. Go thala mebapa ya malaria ka tsela yeo e šomišago maemo a tikologo go na le mehola yeo e oketšegilego kudu go feta ge go balwa batho bao ban ago le malaria maokelong ka gobane tšhomišo ya maemo a tikologo e akaretša le diphetogo tše di tswalanago le maemo a leratadima a nako e telele. Go oketša moo, mafelong ao go ona go sa tsenegego gabonolo goba moo go se nago dipalopalo tša malaria, go thala mebapa ya malaria go šomišwa thekinolotši ya lefaufaung go dira gore re be le kwešišo e tseneletšego ka phatlalalo ya malaria mafelong ao. Morero wa nyakišišo ye e ebe ele go thabala mebapa ya menang yeo e bitšwago *Anopheles arabiensis* le *Anopheles fenestus* Seleteng sa Masepala wa Vhembe ya Afrika Borwa, go dirišwa mekgwa ya disatalaete. Bobedi Landsat TM le Sentinel-2 di dirišitšwe mo nyakišišong ye. Dinyakišišo di bontšha gore go ba gona ga twatši ya *P. falciparum* go tswalana kudu le monola wa mehlare le botala bja yona. Ka go le lengwe, mangatha a dikgomo le o na a tswalana le moo menang e tswalago gona. Nyakišišo ye gape e bontšhitše ka lekga la mathomo, gore go a kgonega go thala mebapa ya *Mošunkwane* e lego wo o dirišwago kudu bakeng sa go thibela menang ka gae. Go hweditšwe gore phatlalalo ya *Mošunkwane* e tswalana kudu le dithaba, tshekamelolo ya thaba, le botala bja dihlare. Se se bopa karolo e bohlokwa bakeng sa maiteko a go fediša malaria Seleteng sa Vhembe.

## List of Figures

	<b>Contents</b>	<b>Page</b>
<b>Figure 1.1</b>	The location of the study area in the Limpopo province of South Africa	9
<b>Figure 1.2</b>	Diagrammatical representation of thesis outline adopted in this thesis	10
<b>Figure 2.1</b>	The average spectral reflectance of the three environmental features that affect malaria transmission (water, soil and green vegetation).	23
<b>Figure 3.1</b>	Summary of the methodology adopted for the study	59
<b>Figure 3.2</b>	Predicted cattle hoofprints values (per 100 m <sup>2</sup> ) using (a) ordinary kriging, (b) step-wise multiple linear regression, and (c) co-kriging. The kriging methods were fitted using exponential model.	68
<b>Figure 3.3.</b>	Predicted cattle hoofprints in the Vhembe District Municipality using co-kriging of field data and S-2 spectral data	69
<b>Figure 3.4</b>	A semi-variogram for the co-kriging method using exponential model.	71
<b>Figure 3.5</b>	The recently updated malaria risk map of at the study area.	72
<b>Figure 4.1</b>	The location of the study area	89
<b>Figure 4.2.</b>	The histogram showing the frequency of LAI distribution per category. In this current study, the highest number of field observations had LAI values of between 2 and 4 m <sup>2</sup> /m <sup>-2</sup>	90
<b>Figure 4.3</b>	Predicted vs. measured LAI using three linear models (a) BBVI model, (b) NBVI model and (c) final model.	91
<b>Figure 4.4</b>	The results of LAI retrieval from narrow-band vegetation indices (a) vs. the final broadband vegetation indices (b) at the study area. The scatterplots show the respective correlation between measured and predicted LAI for the NBVI (c) and the final BBVI model (d) respectively.	96
<b>Figure 5.1.</b>	The study area in the northern part of Limpopo. Each coloured polygon represents individual villages under consideration ( <i>n</i> = 28).	113
<b>Figure 5.2</b>	Overview of epidemiological and remote sensing datasets and methods used for the study	116



<b>Figure 5.3</b>	Representation of different buffer distances across 28 villages in Vhembe District Municipality. Buffer distances were ranging from 0.5km to 20 km.	117
<b>Figure 5.4</b>	Logistic regression performance across threshold values of 0.5-1.0 as applied on buffer distances selected for the study area using validation dataset ( $n=34$ )	122
<b>Figure 5.5</b>	The $D^2$ and AIC of logistic regression applied on different buffer distances in VDM	123
<b>Figure 5.6</b>	Residual plots of environmental covariates used for mapping malaria distribution in VDM.	124
<b>Figure 5.7</b>	Predicted potential geographic distribution of malaria produced through logistic regression and Landsat-derived environmental covariates. Maps produced at buffer distances of 0.5 km (a), 1km (b), 5 km (c), 10 km (d) and 20 km (e).	126
<b>Figure 5.8</b>	Predicted spatial distribution of malaria in VDM at a probability threshold of $\geq 0.7$ . Red colour depicts areas of predicted <i>P. falciparum</i> presence, while white represents predicted <i>P. falciparum</i> absence and Soutpansberg Mountain mask.	131
<b>Figure 6.1</b>	Limits of the study area (Vhembe District Municipality) and the sample locations used for modelling (shown by red dots).	144
<b>Figure 6.2</b>	Analysis of variable importance for both logistic regression (a) and Maxent (b) predictive models ( $n=106$ ).	152
<b>Figure 6.3</b>	ROC curves of different models used in this chapter (a) logistic regression, (b) Maxent and (c) ensemble model. The dotted line depicts a line of no-discrimination (random guess) while the red line indicates sensitivity and specificity at various threshold levels.	153
<b>Figure 6.4</b>	Comparison of predictions for <i>L. javanica</i> using (a) logistic regression model, (b) Maxent and (c) the weighted ensemble modelling. The green colour indicates areas of low probability of occurrence while the red colour corresponds to areas of high species occurrence probability.	154
<b>Figure 6.5</b>	Ecological niche modelling. Map of the predicted distribution showing the percentage of relative habitat suitability for <i>L. javanica</i> species.	158

- Figure 7.1** Map that shows the distribution of cattle hoofprints per 100 m<sup>2</sup> in the Vhembe District Municipality 163
- Figure 7.2** The results of LAI retrieval from narrow-band vegetation indices (a) vs. the final broadband vegetation indices (b) at the study area. The scatterplots show the respective correlation between measured and predicted LAI for the NBVI (c) and the final BBVI model (d) respectively 165
- Figure 7.3** The accuracy of malaria mapping when pseudo-absences are drawn from 0.5, 1, 5, 10 and 20 km from the known presence locations 166



## List of Tables

	<b>Content</b>	<b>Page</b>
<b>Table 2.1</b>	Some of the major globally available EO satellites to date. Satellites are arranged in alphabetical order and not according to importance.	30
<b>Table 3.1</b>	Descriptive statistics of the measured cattle hoofprints in the study area	63
<b>Table 3.2</b>	Multispectral bands of S-2 explored in this chapter	64
<b>Table 3.3</b>	Spectral indices derived from S-2 that were tested in the current study	64
<b>Table 3.4</b>	The relationship between cattle hoofprints and spectral datasets ( $n = 77$ ).	69
<b>Table 3.5</b>	The S-2 variables used in the final SMLR model	70
<b>Table 3.6</b>	Comparison of interpolation performance among OK, CK and SMLR for predicting cattle hoofprints	70
<b>Table 4.1</b>	Descriptive statistics of the measured field LAI in the current study	94
<b>Table 4.2</b>	The vegetation indices used in the study, utilizing green, red NIR and red edge spectral regions	96
<b>Table 4.3</b>	The results of the initial combined BBVI and NBVI regression model and the variable correlations to measured LAI	98
<b>Table 4.4</b>	The results of the significant BBVI model used for mapping LAI ( $n = 66$ )	96
<b>Table 5.1</b>	Selected remote sensing indices that were employed for the study. All indices are derived from Landsat TM	124
<b>Table 5.2</b>	Results of the logistic regression along 5 buffer distances from the known <i>P. falciparum</i> locations.	127
<b>Table 6.1</b>	Broad-band and narrow-band spectral indices used in modelling the distribution of <i>L. javanica</i> at the study area.	156





<b>Table 6.2</b>	Descriptive statistics of S2 and SRTM-derived variables at presence and absence sites for <i>Lippia javanica</i> species.	160
<b>Table 6.3</b>	Correlations between satellite data and the <i>L. javanica</i> presence/absence data ( $n = 151$ )	160
<b>Table 6.4:</b>	The final predictive model selected using stepwise binary logistic regression	161
<b>Table 6.5</b>	Predicted occurrence vs. observed occurrence of <i>L. javanica</i> ( $n = 45$ ) using three SDM's at probability threshold $> 0.6$ .	165
<b>Table 7.1</b>	Comparison of interpolation performance among OK, CK and SMLR for predicting cattle hoofprints	189
<b>Table 7.2</b>	Predicted occurrence vs. observed occurrence of <i>L. javanica</i> ( $n = 45$ ) using three SDM's at probability threshold $> 0.6$ .	194



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

*To my two beloved sons, Thabang and Thato*



## Table of Contents

Declaration.....	iii
Disclaimer .....	iv
Acknowledgments.....	v
Publications and Manuscripts.....	vi
Kakaretšo.....	viii
List of Figures .....	ix
List of Tables .....	xii
CHAPTER 1:.....	1
General Introduction .....	1
1.1 Introduction .....	2
1.2 Remote sensing of infectious diseases .....	2
1.3 Remote sensing for soil, water and vegetation.....	3
1.4 Statistical methods for analysis of soil, water and vegetation .....	5
1.5 Motivation of the study .....	5
1.6 Aim .....	7
1.7 Objectives.....	7
1.8 Scope of the study .....	7
1.8.1 Study area .....	7
1.8.2 Thesis outline.....	10
1.9 References .....	12
CHAPTER 2:.....	17
Literature review .....	17
Abstract.....	18
2.1 Introduction .....	18
2.2 Malaria vector habitats.....	21
2.3 Reflectance properties of water.....	22
2.4 Reflectance properties of soil.....	23



2.5	Reflectance properties of vegetation .....	25
2.6	Mapping malaria: remote sensing perspective .....	26
2.7	Malaria mapping: data challenges.....	32
2.8	Methodological challenges.....	33
2.8.1	Classification.....	33
2.8.2	Regression .....	33
2.9	Opportunities .....	34
2.10	Further research requirements in remote sensing for malaria mapping.....	35
2.11	Conclusions .....	36
2.12	References .....	37
CHAPTER 3:	.....	53
	Integrating geostatistics and remote sensing for mapping the spatial distribution of cattle hoofprints in relation to malaria vector control.....	53
	Abstract .....	54
3.1.	Introduction .....	54
3.2.	Methods.....	58
3.2.1.	Study area and malaria data .....	58
3.2.2.	Field data acquisition and transformation.....	58
3.2.3.	Remote sensing data acquisition and pre-processing .....	60
3.2.4.	Statistical analysis .....	62
3.2.5.	Calibration model.....	65
3.2.6.	Validation model .....	65
3.3.	Results.....	65
3.3.1.	Relationship between cattle hoofprints and S-2 data.....	65
3.3.2.	Spatial interpolation of cattle hoofprints.....	66
3.3.3.	Mapping of cattle hoofprints .....	69
3.4.	Discussion .....	69
3.5.	Conclusions .....	73



3.6. References .....	73
CHAPTER 4:.....	85
Comparison of the Sentinel-2 broadband and narrowband vegetation indices for mapping LAI in malaria-prone heterogeneous semi-arid environment of Southern Africa .....	85
Abstract .....	86
4.1. Introduction .....	86
4.2. Methods.....	89
4.2.1. Study area .....	89
4.2.2. Field data.....	90
4.3.3. Image data.....	91
5.2.1. Statistical analysis .....	92
4.3. Results.....	93
4.3.1. Broadband and narrow-band models .....	93
5.3.1. Accuracy assessment.....	95
4.4. Discussion .....	95
5.4.1 Spatial mapping and considerations .....	98
4.5. Conclusions .....	98
4.6. References .....	99
CHAPTER 5:.....	109
Evaluating Efficacy of Landsat-Derived Environmental Covariates for Predicting Malaria Distribution in Rural Villages of Vhembe District, South Africa .....	109
Abstract .....	110
5.1. Introduction.....	110
5.2. Methods.....	113
5.2.1. Study area.....	113
5.2.2. Epidemiological data .....	114
5.2.3. Remote sensing data .....	115
5.2.4. Data pre-processing .....	116



5.2.5. Data processing and analysis.....	118
5.2.6. Remote sensing data analysis.....	118
5.2.7. Statistical analysis.....	120
5.3. Results.....	122
5.3.1. Predictive maps.....	126
5.4. Discussions.....	126
5.4.1. Remote sensing environmental covariates .....	128
5.4.2. Predictive maps.....	130
5.4.3. Limitations and recommendations .....	133
5.5. Conclusions .....	133
5.6. References .....	134
CHAPTER 6:.....	141
Mapping the spatial distribution of <i>Lippia javanica</i> (Burm. f.) Spreng using Sentinel-2 and SRTM-derived topographic data in malaria endemic environment .....	141
Abstract .....	142
6.1 Introduction .....	142
6.2 Materials and methods.....	146
6.2.1 Study area .....	146
6.2.2 Field data.....	147
6.2.3 Remote sensing/topographical data.....	148
6.2.4 Data pre-processing.....	148
6.2.5 Modelling strategy.....	150
6.2.6 Model evaluation.....	152
6.3 Results.....	152
6.3.1 Predictive maps .....	155
6.1 Discussion .....	158
6.2 Conclusions .....	162
6.3 References .....	163



CHAPTER 7:.....	177
Remote sensing of environmental variables for mapping malaria distribution: the case of Vhembe District Municipality, South Africa.....	177
Summary and conclusions.....	177
7.1 Introduction.....	178
7.1.1 Challenges and opportunities of remote sensing for malaria mapping .....	179
7.1.2 Estimating the distribution of cattle hoofprints .....	180
7.1.3 Characterizing mosquito resting and questing habitats (LAI).....	182
7.1.4 Mapping the distribution of malaria in the study area .....	184
7.1.5 Mapping the spatial distribution of <i>L. javanica</i> for malaria control.....	185
7.2 Overall conclusions and future work .....	187
7.3 Abridged Summary .....	189
7.4 References .....	190
Curriculum Vitae .....	194



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

Page intentionally left blank





UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

# **CHAPTER 1:**

## General Introduction

This chapter provides the background of the study and outlines the aim, objectives and structure of this thesis

## 1.1 Introduction

This chapter provides brief introduction to the study by highlighting the historical perspectives on the use of geospatial technology for infectious diseases mapping including malaria. It demonstrates how remote sensing data are associated with important environmental conditions influencing malaria transmission in subtropical environment. Furthermore, the remote sensing methods commonly used for assessing vegetation, water and soil conditions in relation to malaria vector habitats are discussed. The major components of malaria vector habitats discussed in this study are (i) cattle hoofprints, and the (ii) vegetation characteristics (leaf area index, greenness). The chapter also provides the aim, objectives and the outline of the study.

## 1.2 Remote sensing of infectious diseases

Throughout centuries, health professionals have relied on spatial information in the form of maps to derive important baseline information on disease transmission (Guerra *et al.*, 2008) and as a basis for resource allocation at various scales (Pigott *et al.*, 2015). Some of the earliest infectious diseases mapped include cholera epidemic mapping that took place in 1831 and spanned over several decades (Gilbert, 1958). The initial objective was to identify the hotspots within densely population urban settlements which lacked property sanitation (Musa *et al.*, 2013). However, it was only in 1931-33 when Petermann correlated the infectious disease incidences with certain environmental conditions affecting disease transmission (Gilbert, 1958). In Africa, some efforts have been done to map malaria and other infectious diseases by linking environmental factors peculiar to the habitats of disease vectors. For example, in 1997 Thomson *et al.* (1997) successfully demonstrated the contribution of remote sensing variables to characterizing malaria vector habitat in The Gambia. In South Africa, the use of geospatial technology for malaria mapping was initially realized in 1990 when the GIS maps were produced for two districts in the Northern KwaZulu-Natal. This initiative was part of the Malaria Research Programme (MRP), which was subsequently followed by the famous malaria mapping project called Mapping Malaria Risk in Africa/Atlas du Risque de la Malaria en Afrique (MARA/ARMA)(Adeola *et al.*, 2015). When mapping malaria, the environmental variables such as proximity to standing water bodies, urbanization, irrigation, vegetation greenness, and vegetation moisture have shown correlation with the disease incidence rates (Omumbo *et al.*, 2002; Adeola *et al.*, 2015).

For efficient malaria mapping, consideration of such factors will, not only contribute to accurate quantification of disease hotspots but, also the baseline for targeted disease control and early-warning system. Disease maps have been used to represent the manifestations of complex geographical and epidemiological phenomena in more simplistic manner than count data for centuries. They provide key interface between scientists and health professionals regarding status of malaria distribution (including outbreaks), environmental conditions at vector habitats, population at risk, and malaria forecasting (Dalrymple *et al.*, 2015). The conventional approach to mapping malaria patterns involves ground-based surveys, which are labour-intensive and expensive. Additionally, the identification of potential habitats of malaria vectors is conventionally done manually, which can be cumbersome for wide-area surveillance. Fortunately, remote sensing technology offers rapid and efficient alternative approach for mapping malaria distribution in the endemic areas. Some of the greatest advantages of remote sensing technique to traditional malaria mapping include the wide-area coverage, ability to acquire datasets in very rugged and inaccessible areas, high correlation with potential habitats of malaria vectors, and varying spatio-temporal characteristics necessary for diseases surveillance (Adeola *et al.*, 2015; Malahlela *et al.*, 2018). Today, the open-access big data, high-speed processing software and high computational capabilities open new opportunities for fine-scale malaria mapping. These capabilities offer improved procedures for mapping malaria vector habitats related to soil, water and vegetation biophysical and structural characteristics (Dlamini *et al.*, 2015).

Sections 1.3 and 1.4 of this thesis highlight the contribution of remote sensing techniques and the statistical methods used for mapping soil, water and vegetation characteristics, which have shown high correlation to malaria transmission rates in endemic areas. In section 1.5 there is a detailed motivation of the study that is provided to justify the research conducted and reported in this thesis.

### **1.3 Remote sensing for soil, water and vegetation**

Several remote sensing methods have been applied for mapping the biophysical and biochemical characteristics of soil, water and vegetation and such biophysical/chemical characteristics directly and indirectly linked to malaria vector habitats and disease transmission. For example, Midekisa *et al.* (2012) have correlated the actual evapotranspiration (ETa) from the soil and vegetation to malaria transmission over time. In this study, it was found that malaria transmission is positively correlated to the lag in ETa from 1-3

months. Furthermore, [Dlamini et al. \(2015\)](#) alluded that in southern Africa, the small water pools which are related to anthropogenic activities tend to be favourable habitats for *Anopheline* mosquito. It is from this basis that [Hardy et al. \(2017\)](#) mapped water bodies in Zanzibar by using remote sensing techniques. Mapping of water bodies has been extensively done using various remote sensing data acquired at varying spatial and temporal resolutions ([Acharya et al., 2018](#); [Liu et al., 2017](#); [Malahlela et al., 2016](#)). In addition, mapping of water quality through remote sensing is regarded as one the crucial steps for water conservation, risk assessment and malaria estimations ([Kengluetcha et al., 2005](#)). However, although mapping the spatial extent of water bodies has proven successful through various remote sensing approaches, spatial modelling of micro-habitats of malaria vector such as cattle hoofprints is currently not available at any scale. This is despite global knowledge regarding the importance of cattle hoofprints and puddles in malaria transmission ([Mayagaya et al., 2015](#); [Munga et al., 2007](#)). The micro-habitats, such as puddles, water pipes, artificial containers, cattle hoofprints, neglected wells and artificial ditches have been documented as some of the crucial breeding sites for *Anopheles arabiensis* mosquito ([Hamza and El Rayah, 2016](#)). The occurrence of micro-habitats is very crucial for malaria vector, particularly considering that macro-habitats (such as dams, lakes and rivers) comprise of various macro-invertebrates (e.g. fish, amphibians, dragonflies) that prey on mosquito larvae ([Mogi, 2007](#)). Therefore, the spatial information pertaining to the distribution of malaria vector micro-habitats is crucial for targeted outdoor malaria control and management.

Malaria transmission has shown high correlation with vegetation greenness in many tropical areas ([Dlamini et al., 2015](#); [Ricotta et al., 2014](#); [Adeola et al., 2016](#)). For example, [Ricotta et al. \(2014\)](#) observed that vegetation cover found at close proximity to homestead in East Zambia is usually associated with high number of malaria vector that rest in nearby vegetation of varying types and densities. Malaria vectors such as *Anopheles arabiensis*, *An.fenestus* and other species of *An.* complex prefer to rest and quest under vegetation leaves in order to avoid predators and heat stress that could result in desiccation of individual malaria vector ([Dewald et al., 2016](#); [Paaijmans and Thomas, 2011](#)). This means that the canopy architecture and the plant leaves play a significant role for the *Anopheline* species, and the understanding of this biophysical parameter could enhance the understanding of malaria distribution in tropical environments. The high resolution remote sensing data, such as Sentinel-2 and commercially available WorldView-2/3, open new opportunities for accurate characterization of plant canopy and leaf area. Combining these new generation remote sensing datasets with

statistical methods enhances our understanding of the environments suitable for malaria vector survival and reproduction.

#### **1.4 Statistical methods for analysis of soil, water and vegetation**

In remote sensing, many various statistical methods that combine information from the physical environment and spectral bands are increasingly being utilized over the years. One of the commonly used approaches includes multivariate regression models, which relate the biophysical and biochemical parameters of soil, water and vegetation with reflectance or absorbance of electromagnetic radiation in linear form (Hudak *et al.*, 2006; de Jong and de Bruin, 2012; Xavier and Vettorazzi, 2004; Nyamugama and Kakembo, 2014; Gomez *et al.*, 2008; Malahlela *et al.*, 2018). For example, Forkuor *et al.* (2017) compared the use of multiple linear regression model with other statistical models to map soil properties by means of remote sensing variables in the South-Western Burkina Faso.

Another form of generalized linear model called logistic regression model has extensively been used for mapping soil (Kempen *et al.*, 2009), water (Mueller *et al.*, 2016), and vegetation characteristics (Aspinall, 2002). Most commonly, the logistic regression methods has been used in species distribution mapping because of the ability to handle the interactions of various environmental variables with the binary nature of species occurrence. On the other hand, the maximum entropy model (Maxent) is increasingly receiving more attention as an alternative method for species distribution mapping. For example, Dudov (2017) modelled the distribution of sixty-three common vascular plant species of the Zeya Reserve using Maxent and remote sensing data. It is, however, noticed that both the logistic regression model and Maxent have their short-comings, which limit their individual application as species distribution model. In order to circumvent these limitations, the use of ensemble model (which combines the predictive powers of both models) is recommended and is thus viewed as an alternative to single-model approach (Zhang *et al.*, 2015).

#### **1.5 Motivation of the study**

The World Health Organization (WHO) has proposed the interruption of indigenous approach to malaria transmission in specific geographical areas as a means to combat ever increasing transmission rates in Africa. Similarly, South Africa is one of the countries with planned elimination by the year 2020 (WHO, 2018). For effective implementation of malaria elimination and control programs in South Africa, it thus becomes crucial to have enhanced understanding on the spatial distribution of malaria and environmental conditions associated with high

incidence rates ([Machault et al., 2012](#)). The cases of countries which successfully achieved zero malaria transmission are a compelling reason why geospatial technologies such as remote sensing and geographic information system (GIS) are essential for malaria elimination within the multi-disciplinary framework ([Giraudoux et al., 2008](#)). In South Africa, fortunately, the success of the geospatial technology has been documented by several researchers which has significantly contributed to achieving the reduced malaria transmission rates in malaria-prone areas ([Moonasar et al., 2016](#); [Adeola et al., 2016](#); [Booman et al., 2003](#); [Martin et al., 2002](#)). Such success is also attributed to a wide range of geospatial datasets including point, raster and non-point datasets that are available at high spatial resolution.

The advent of new generation satellite data opens new avenues for mapping malaria at spatial, temporal and spectral resolutions that are suitable for dynamic nature of malaria transmission. The launch of earth observation satellites with open-access datasets enables rapid and efficient analysis of environmental conditions that are associated with malaria transmission, thus playing a critical role as inputs within the malaria information system (MIS). For example, [Adeola et al. \(2016\)](#) utilized Landsat-derived land use/land cover for determining the population at risk of malaria in the Nkomazi municipality in South Africa. The environmental variables such as vegetation greenness and moisture derived from Landsat dataset have been linked with areas of higher malaria incidences in the Vhembe District Municipality ([Malahlela et al., 2018](#)). The provision of these kinds of datasets could provide reliable and accurate information necessary for healthcare workers, policy and decision-makers, as well as the community members for informed malaria intervention strategies and control.

Deriving environmental variables with high mapping accuracies is fundamental for detailed understanding of factors that have influence on the transmission patterns of malaria particularly in remote areas. In light with the geospatial requirements (high temporal and spatial resolution), this study contributes to the enhanced understanding of the environmental variables that promote malaria transmission in subtropical areas of South Africa in order to support efforts aimed at malaria elimination and control. It demonstrates methods for mapping the habitats of malaria vector using new high resolution remote sensing data, and how such datasets contribute to malaria vector control through the ethnobotanical use of plant species by the local community



## 1.6 Aim

The aim of this research was to investigate the potential use of remote sensing-derived environmental variables for mapping malaria distribution in the Vhembe District Municipality, South Africa. In addition to the aim of the study, some specific objectives were set out in the beginning of the study in order to achieve the primary goal of the study, and such objectives are listed in section 1.7 of this general introduction.

## 1.7 Objectives

A total of four (4) objectives are stated below:

- (i) Mapping mosquito micro-habitats using remote sensing technology
- (ii) Estimating the vegetation leaf index (LAI) in relation to known malaria distribution
- (iii) Mapping malaria distribution from the satellite-derived environmental variables, and
- (iv) Predicting the occurrence of *Lippia javanica* (lemon bush) commonly used for repelling malaria in villages located at the study area.

## 1.8 Scope of the study

This thesis investigates the potential use of optical remote sensing data for mapping the distribution of malaria in the study area. The major focus is placed on mapping *Anopheles* larval habitats (cattle hoofprints), adult questing sites (vegetated environments) and the remote sensing application for preventative medicine (lemon bush mapping). This approach was adopted because of the inter-link between various factors involved in malaria transmission, which are mainly environmental-based, and the need to strengthen the sustainable ethnobotanical use of extracts (aroma) for malaria control (Mabogo, 1990; Maroyi, 2017).

### 1.8.1 Study area

The research was conducted in the Vhembe District Municipality of South Africa (VDM), which is situated in the northern-most part of the country between 23°40' S and 30°00'E (Figure 1.1). It comprises of varying topography, with diverse floral and faunal biodiversity. The Municipality receives annual summer rainfall of 820 mm (Mpandeli, 2014), with Soutpansberg Mountain modifying geographical rainfall patterns (Kabanda and Munyati, 2010). The north-western part of Vhembe District is characterized by semi-arid conditions, while the south-eastern part experiences subtropical conditions. The Vhembe District Municipality has a population of more than 1.3 million people (StatSA, 2016), who predominantly reside in rural villages. Data from 28 villages were used for the study. The area covering these villages has recorded mean

malaria incidence of about 328.2 between 1998-1999 and 2004-2005 ([Gerritsen et al., 2008](#)). In 2000, the municipality experienced floods brought by the tropical cyclone *Eline* which dramatically increased malaria cases in Limpopo province ([Reason and Keibel, 2004](#)). Malaria in Limpopo is seasonal and it is, therefore, crucial to use seasonal data covering the entire study area to map the occurrence of malaria in the VDM.



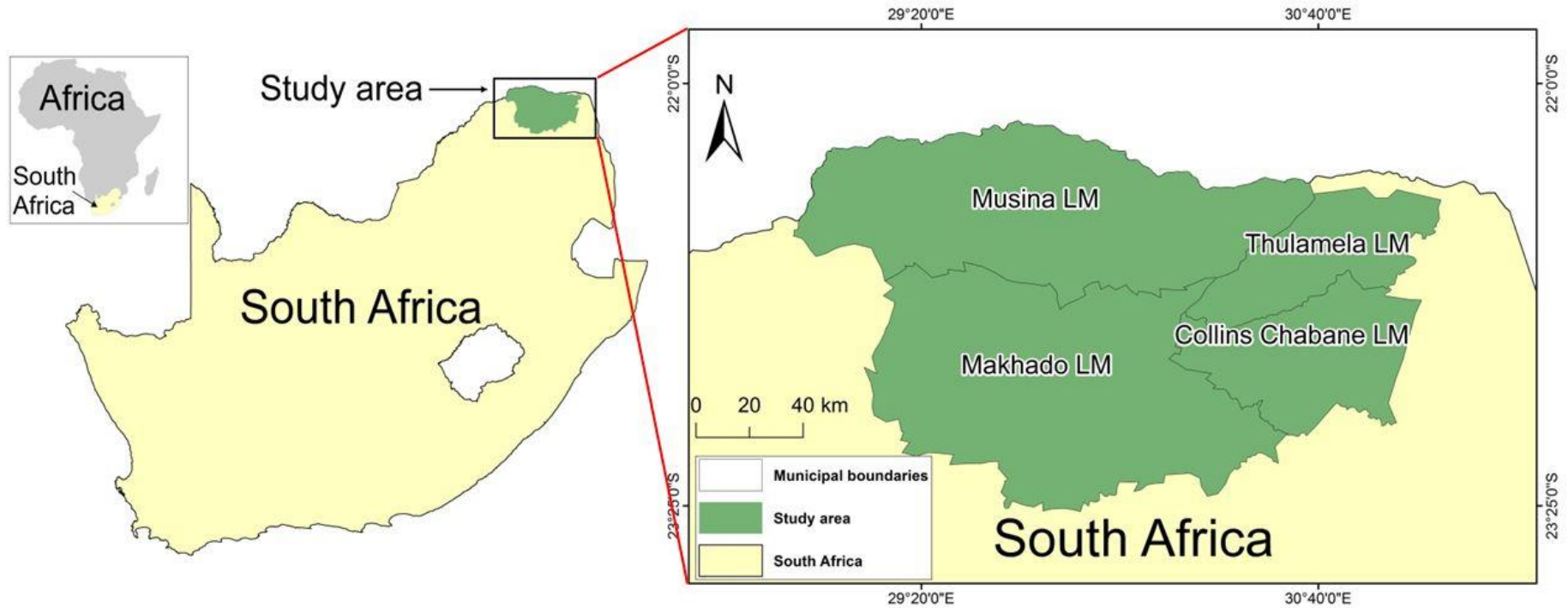
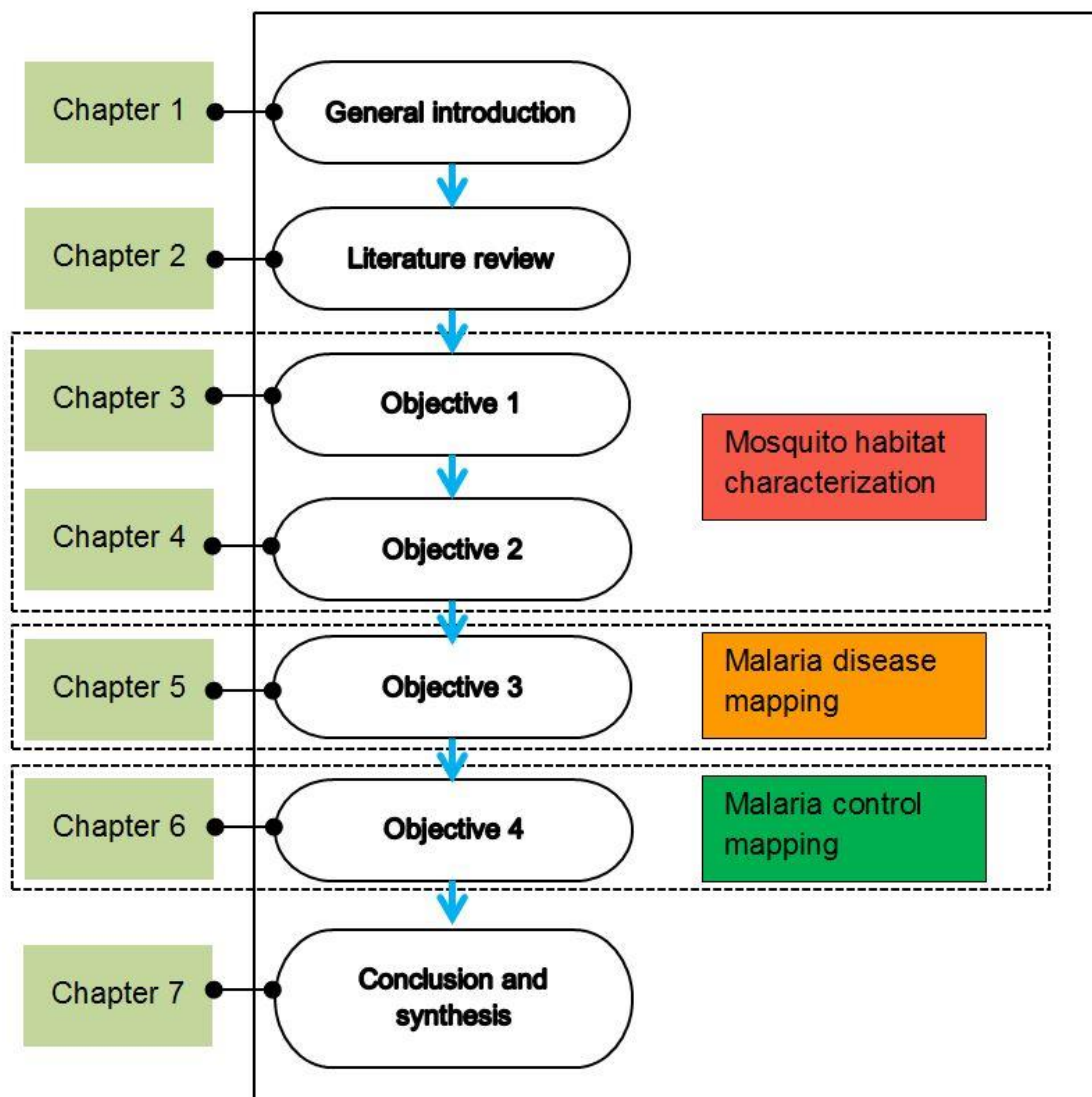


Figure 1.1 The location of the study area in the Limpopo province of South Africa.

### 1.8.2 Thesis outline

This thesis is comprised of collection of 5 papers published and /or to be submitted for peer reviewed in international and local journals. Currently, 3 papers are published in these journals: EcoHealth, Ecological Modelling and International Journal of Remote Sensing, while 2 papers are still in preparation for submission. Each paper is presented separately, as a chapter (Figure 1.2), and is prepared as a final manuscript. The thesis follows a paper format, with each chapter (section) presented as a stand-alone contribution. Each chapter has its own introduction, methods, results, discussion, conclusions, and its own reference list.



**Figure 1.2.** Diagrammatical representation of thesis outline adopted in this thesis



## **Chapter 1**

This chapter provides the overview of the research background, aim and objectives and the structure of the thesis.

## **Chapter 2**

This chapter provides a comprehensive review of remote sensing applications for malaria mapping. It focuses on remote sensing methods, earth observation satellite portfolio, limitations and opportunities available in remote sensing for malaria applications.

## **Chapter 3**

In this chapter the characterization of cattle hoofprints using remote sensing technology is presented. Sentinel-2 dataset is used in this chapter together with geostatistical approaches for mapping the distribution of these important malaria vector habitats.

## **Chapter 4**

This chapter presents mapping of leaf area index (LAI) using high resolution remote sensing dataset at the study area. The LAI is estimated from the spectral indices with high statistical correlation to this vegetation biophysical characteristic.

## **Chapter 5**

This chapter provides mapping of malaria distribution in the Vhembe District Municipality, using environmental variables that are derived from satellite datasets.

## **Chapter 6**

In this chapter, the distribution of *Lippia javanica* (lemon bush) species is mapped. This plant is commonly used in the study area as an ethnobotanical plant. The mapping of *L. javanica* formed an important part of the current thesis especially that this was the initial mapping of this plant species at high spatial and spectral resolution.

## **Chapter 7**

This chapter provides synthesis of this research and highlights the opportunities for future research.

## 1.9 References

- Achrya T., Subedi A., Lee D. (2018). Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. *Sensors*, 18(8):2580.
- Adeola M.A., Botai J.O., Olwoch J.M., de W Rautenbach H.C.J., Kalumba A.M., Tsela P.L., Adisa M.O., Wasswa N.F., Mmtoni P., Ssentongo A. (2015). Application of geographical information system and remote sensing in malaria research and control in South Africa: a review. *South African Journal of Infectious Disease*, 30(4): 7-14.
- Adeola A.M., Botai O.J., Olwoch J.M., Rautenbach C.J., Adisa O.M., Taiwo O.J., Kalumba A.M. (2016). Environmental factors and population at risk of malaria in Nkomazi municipality, South Africa. *Tropical Medicine and International Health*, 21(5): 675 – 686.
- Aspinnall R.J. (2002). Use of logistic regression for validation of maps of the spatial distribution of vegetation species derived from high spatial resolution hyperspectral remotely sensed data. *Ecological Modelling*, 157(2/3): 301 – 312.
- Booman M., Sharp B.L., Martin C.L., Manjate B., la Grange J.J., Durrheim D.N. (2003). Enhancing malaria control using a computerised management system in southern Africa. *Malaria Journal*, 2:13:DOI: 10.1186/1475-2875-2-13.
- Dalrymple U., Mappin B., Gething P.W. (2015). Malaria mapping: understanding the global endemicity of falciparum and vivax malaria. *BMC Medicine*, 13:140. DOI: 10.1186/s12916-015-0372-x.
- De Jong R., de Bruin S. (2012). Linear trends in seasonal vegetation time series and the modifiable temporal unit problem. *Biogeosciences*, 9: 71 – 77.
- Dewald J.R., Fuller D.O., Muller G.C., Beier J.C. (2016). A novel method for mapping village-scale outdoor resting microhabitats of the primary African malaria vector, *Anopheles gambiae*. *Malaria Journal*, 15:489. DOI 10.1186/s12936-016-1534-9.
- Dlamini S.N., Franke J., 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(1): DOI: 10.4081/gh.2015.302.
- Dudov S.V. (2017). Modelling of species distribution with the use of topography and remote sensing data on the example of vascular plants of the Tukuringra Ridge low mountain belt (Zeya State Nature Reserve, Amur Oblast). *Biology Bulletin Reviews*, 7(3): 246 – 257.
- Forkuor G., Hounkpatin O.K.L., Welp G., Thiel M. (2017). High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A

Comparison of Machine Learning and Multiple Linear Regression Models. *PLOS ONE*, 12(1): e0170478. DOI: 10.1371/journal.pone.0170478.

- Gerritsen A.A.M., Kruger P., Schim van der Loeff M.F., Grobusch M.P. (2008). Malaria incidence in Limpopo Province, South Africa, 1998 – 2007. *Malaria Journal* 7:1 – 8.
- Gilbert E.W. (1958). Pioneer maps of health and disease in England. *Geographical Journal*, 124(2): 172 – 183.
- Giraudoux P., Raoul F., Pleydell D., Craig P.S. (2008). Multidisciplinary studies, systems approaches and parasite eco-epidemiology: something old, something new. *Parasite*, 15: 469 – 476.
- Gomez C., Rossel R.A.V., McBratney A.B. (2008). Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma*, 146(3/4): 403 – 411.
- Guerra CA., Gikandi P.W., Tatem A.J., Noor A.M., Smith D.L., Smith D.L., Hay S.I., Snow R.W. (2008) The limits and intensity of *Plasmodium falciparum* transmission: Implications for malaria control and elimination worldwide. *PLOS Medicine*, 5: 300–311.
- Hamza A.M., El Rayah E.A. (2016). A Qualitative Evidence of the Breeding Sites of *Anopheles arabiensis* Patton (Diptera: Culicidae) in and Around Kassala Town, Eastern Sudan. *International Journal of Insect Science*, 8: 65 – 70.
- Hardy A., Makame M., Cross D., Majambere S., Msellem M. (2017). Using low-cost drones to map malaria vector habitats. *Parasites and Vectors*, 10(29): DO: 10.1186/s13071-017-1973-3.
- Hudak A.T., Crookston N.L., Evans J.S., Falkowski M.J., Smith A.M.S., Gessler P.E., Morgan P. (2006). Regression modelling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. *Canadian Journal of Remote Sensing*, 32(2): 126 – 138.
- Kabanda, T. and Munyati, C. (2010). Anthropogenic-induced climate change and the resulting tendency to land conflict; The case of the Soutpansberg region, South Africa; Climate Change and Natural Resources Conflicts in Africa, Eds. Donald Anthony Mwiturubani and Jo-Ansie van Wyk. *Monograph No 170: Institute for Security Studies*.
- Kempen B., Brus D.J., Heuvelink G.B.M., Stoorvogel J.J. (2009). Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151(3/4): 311 – 326.

- Kengluetcha A., Singhasivanon P., Tiensuwan M., Jones J.W., Sithiprasansa R. (2005). Water quality and breeding habitats of *Anophele* mosquito in northwestern Thailand. *South-East Asian Journal of Tropical Medicine and Public Health*, 36(1): 46 – 53.
- Liu Y., Wang X., Xu S., Wang C. (2017). Analysis of Coastline Extraction from Landsat-8 OLI Imagery. *Water*, 9(11):816.
- Mabogo D.E.N. (1990). The Ethnobotany of the Vhavenda. M.Sc. Thesis. University of Pretoria.
- Machault V., Vignolles C., Pages F., Gadiaga L., Toure Y.M., Gaye A., Solhna C., Trape J., Lacaux J., Rogier C. (2012). Risk Mapping of *Anopheles gambiae* s.l. Densities Using Remotely-Sensed Environmental and Meteorological Data in an Urban Area: Dakar, Senegal. *PLOS ONE*, 7(11): e50674. DOI: 10.1371/journal.pone.005067.
- Malahlela O.E. (2016). Inland waterbody mapping: towards improving the discrimination and extraction of surface water features. *International Journal of Remote Sensing*, 37(19): 4574 – 4589.
- Malahlela O.E., Oliphant T., Tsoeleng L.T., Mhangara P. (2018). Mapping chlorophyll-a concentrations in a cyanobacteria and algae-impacted Vaal Dam using Landsat 8 OLI data. *South African Journal of Science*, 114(9/10), Art. #4841: DOI: <http://dx.doi.org/10.17159/sajs.2018/4841>
- Malahlela O.E., Olwoch J.M., Adjorlolo C. (2018). Evaluating Efficacy of Landsat-Derived Environmental Covariates for Predicting Malaria Distribution in Rural Villages of Vhembe District, South Africa. *EcoHealth*, 15(1): 23 – 40.
- Maroyi A. (2017). *Lippia javanica* (Burm.f.) Spreng.: Traditional and Commercial Uses and Phytochemical and Pharmacological Significance in the African and Indian Subcontinent. *Evidence-Based Complementary and Alternative Medicine*, 2017: ID.6746071.
- Martin C., Curtis B., Fraser C., Sharp B. (2002). The use of GIS-based malaria information system for malaria research and control in South Africa. *Health and Place*, 8: 227 – 236.
- Mayagaya V.S., Nkwengulila G., Lyimo I.N., Kihonda J., Mtambala H., Ngonyani T., Russell T.L., Furguson H.M. (2015). The impact of livestock on the abundance, resting behaviour and sporozoite rate of malaria vectors in southern Tanzania. *Malaria Journal*, 14(17): DOI: 10.1186/s12936-014-0536-8.





- Midekisa A., Senay G., Henebry G.M., Semuniguse P., Wimberly M.C. (2012). Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. *Malaria Journal*, 11:165. DOI: 10.1186/1475-2875-11-165.
- Mogi M. (2007). Insects and other invertebrate predators. *Journal of American Mosquito Control Association*, 23(suppl 3): 93 – 109.
- Moonasar D., Maharaj R., Kunene S., Candrinho B., Saute F., Ntshalintshali N., Morris N. (2016). Towards malaria elimination in the MOSASWA (Mozambique, South Africa and Swaziland) region. *Malaria Journal*, 15:419. DOI: 10.1186/s12936-016-1470-8.
- Mpandeli S. (2014). Managing Climate Risks Using Seasonal Climate Forecast Information in Vhembe District in Limpopo Province, South Africa. *Journal of Sustainable Development* 7 (5): 68 – 81.
- Mueller N., Lewis A., Roberts D., Ring S., Melrose R., Sixsmith J., Lymburner L., McIntyre A., Tan P., Curnow S., Ip A. (2016). Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sensing of Environment*, 174: 341 – 352.
- Munga S., Minakawa N., Zhou G., Githeko A.K., Yan G. (2007). Survivorship of Immature Stages of *Anopheles gambiae* s.l. (Diptera: Culicidae) in Natural Habitats in Western Kenya Highlands. *Journal of Medical Entomology*, 44(5): 758 – 764.
- Musa G.J., Chiang P., Sylk T., Bavley R., Keating W., Lakew B., Tsou H., Hoven C.W. (2013). Use of GIS Mapping as a Public Health Tool—From Cholera to Cancer. *Health Services Insights*, 6: 111 – 116.
- Nyamugama A., Kakembo V. (2014). Estimation and monitoring of aboveground carbon stocks using spatial technology. *South African Journal of Science*, 111(9/10): Art.#2014-0170.
- Omumbo J.A., Hay S.I., Goetz S.J., Snow R.W., Rogers D.J. (2002). Updating Historical Maps of Malaria Transmission Intensity in East Africa Using Remote Sensing. *Photogrammetry Engineering and Remote Sensing*, 68(2): 161 – 166.
- Paaijmans K.P., Thomas M.B. (2011). The influence of mosquito resting behaviour and associated microclimate for malaria risk. *Malaria Journal*, 10:183. DOI: 10.1186/1475-2875-10-183.
- Pigott D.M., Howes R.E., Wiebe A., Battle K.E., Golding N., Gething P.W., Dowell S.F., Farag T.H., Garcia A.J., Kimball A.M., Krause L.K., Smith C.H., Brooker S.J., Kyu H.H., Vos T., Murray C.J.L., Moyes C.L., Hay S.I. (2015). Prioritizing infectious disease mapping. *PLOS Neglected Tropical Diseases*, 9(6):e0003756. DOI: 10.1371/journal.pntd.0003756.



- Reason C.J.C., Keibel A. (2004). Tropical cyclone *Eline* and its unusual penetration and impacts over the southern African mainland. *Weather Forecast*, 19(5):789–805.
- Ricotta E.E., Frese S.A., Choobwe C., Louis T.A., Shiff C.J. (2014). Evaluating local vegetation cover as a risk factor for malaria transmission: a new analytical approach using ImageJ. *Malaria Journal*, 13(94): DOI: 10.1186/1475-2875-13-94.
- Statistics South Africa. (2016). Population census 2016. Vhembe District Municipality. [http://cs2016.statssa.gov.za/?page\\_id=270](http://cs2016.statssa.gov.za/?page_id=270)
- World Health Organization (WHO). (2018). Update on the E-2020 Initiative of 21 Malaria-Elimination Countries: *Report and Country Briefs*. WHO Geneva, Switzerland.
- Xavier A.C., Vettorazzi A.C.(2004). Mapping leaf area index through spectral vegetation indices in a subtropical watershed. *International Journal of Remote Sensing*, 25(9): 1661 – 1672.
- Thomson M.C., Connor S.J., Milligan P., Flasse S.P. (1997). Mapping malaria risk in Africa: what can satellite data contribute? *Parasitology Today* 13(8): 313 – 318.
- Zhang L., Liu S., Sun P., Wang T., Wang G., Zhang X., Wang L. (2015). Consensus Forecasting of Species Distributions: The Effects of Niche Model Performance and Niche Properties. *PLOS ONE*, 10(3): e0120056. DOI: 10.1371/journal.pone.0120056.





## CHAPTER 2<sup>1</sup>: Literature review

---

<sup>1</sup>This chapter is based on the manuscript titled “*Applications of remote sensing for malaria mapping: challenges and opportunities*” (In preparation)

## Abstract

Globally, malaria continues to affect around 91 countries with over 216 million cases reported in 2016. The malaria disease exerts huge burden on socio-economic infrastructures and investments of the countries directly or indirectly affected by the parasite. Many efforts were made towards supporting the elimination of malaria – a target set out by the World Health Organization. One such effort involves mapping of habitats suitable for malaria vectors such as the *Anopheles* mosquito complex using geospatial techniques. Remote sensing techniques are increasingly being recognized as alternative for mapping disease occurrence, forecasting and estimation globally. There is a growing interest in the application of high resolution geospatial datasets for mapping malaria so as to contribute to operational malaria early-warning systems. The major focus of such mapping revolves around identification of potential habitats for *Anopheles* mosquitoes for targeted malaria control and management. In this chapter, we review the applications of remote sensing for malaria mapping. We then highlight various remote sensing data available for malaria vector habitats, challenges relating to remote sensing data and the methods commonly used for mapping malaria. We conclude by providing potential opportunities for further studies as well as the major findings from the literature review.

**Keywords:** malaria; early-warning system; *Anopheles* mosquitoes; habitats

---

## 2.1 Introduction

Throughout the centuries, spatial modelling of infectious diseases such as malaria has become a fundamental requirement for effective disease control, management and elimination. Over the years, malaria mapping continued to receive more attention and such efforts were epitomized mainly by two main initiatives: the (i) Global Burden of Diseases, and (ii) the Disease Control Priorities (Lopez *et al.*, 2006; Jamison *et al.*, 2006; Hay and Snow, 2006). The largest effort ever realized for mapping malaria was done for mapping *Plasmodium falciparum* spatial occurrence in Kenya in the 1960s, and has since been cited as an attempt to map malaria at a global scale (Hay and Snow, 2006). This attempt was initiated by Lysenko and Semashko (1968), representing a large synthesis of historical datasets including parasite rate, entomological inoculation rate and malaria vector distributions. These maps laid a ground work with regards to the developments that would soon be realized in the 20<sup>th</sup> century going forward.

In nearly half a century since Lysenko and Semashko (1968) produced the global malaria maps, a major initiative encompassing continent-wide mapping of malaria disease at high resolution emerged. This initiative was dubbed Mapping Malaria Risk in Africa/Atlas du Risque

de la Malaria en Afrique (MARA/ARMA) project. This project was instigated in 1997, nearly 30 years since the first global malaria map production. The MARA/ARMA project has transformed the arts, science, economics and modelling of malaria disease in the modern history. The success of MARA/ARMA, led by David LeSeur, spanned from comprehensive collection of African malariometric datasets, development of base-maps relating to malaria risk in Africa, and to determine the environmental/climatic factors that either promote or demote the likelihood of malaria in African continent. In this project, mapping was predominantly done using geospatial statistics in geographic information systems (GIS). These contributions were highly recognized and cited by many researchers including [Dalrymple \*et al.\* \(2015\)](#), [Adeola \*et al.\* \(2016\)](#), [Snow and Noor \(2015\)](#) and [Thomas \*et al.\* \(2004\)](#). Some of the methods used for mapping malaria distribution included (i) regression modelling ([Adeola \*et al.\*, 2016](#)), (ii) classification ([Mohan and Naumova 2014](#)), (iii) multi-criterion assessments ([Alimi \*et al.\*, 2016](#)), and (iv) machine learning (fuzzy) approaches ([Weiss \*et al.\*, 2015](#)). In these methods, efforts are made to link the environmental factors with the likelihood of malaria occurrence and abundance in a particular area of interest. In this way, mapping of malaria inherently depends on the dynamics environmental/climatic variables, input data and the statistical/mathematical methods used for modelling malaria.

Many of the environmental factors such as distance to water bodies, elevation, distance to nearby settlements, and vegetation greenness were considered important for accurate mapping of malaria, due to their direct or indirect link with malaria vectors ([Adeola \*et al.\*, 2016](#)). Obtaining information concerning these sets of environmental and topographical variables is conventionally done by means of field surveys. The limitations with obtaining data in this manner arise due to the physical and socio-economic difficulties associated with malaria studies and population at risk. For example, collecting environmental variables related to *Anopheles* breeding sites and resting habitats is laborious and time-consuming and is often hampered by the inaccessibility of sampling areas ([Malahlela \*et al.\*, 2018](#)). Collecting data in malaria-prone environments also exposes researchers to potential threat of malaria transmission especially during peak transmission season. This makes the use of field-based methods limited especially in landscapes where access is hampered by physical barriers such as mountain ranges, rivers or in instances where wild animals pose potential threat to both livestock and humans. These limitations call for alternative methods that are capable of providing wider spatial coverage, up-to-date, rapid and accurate information regarding the environmental/climatic and topographical factors associated with malaria mapping.

Satellite remote sensing technology provides rapid and efficient alternative for mapping malaria. Studies that utilize remote sensing for mapping malaria typically relate malaria cases with environmental factors such as vegetation greenness and water bodies whose presence and intensity is subject to seasonal variations (Sarmah *et al.*, 2018; Kaptué *et al.*, 2013). These correlations are thus important in malaria prediction in that vegetation greenness usually serve as a *proxy* for mosquito presence, while the water bodies typically show a positive correlation with the larval densities in many areas (Valle *et al.*, 2013; Soleimani-Ahmadi *et al.*, 2014; Amadi *et al.*, 2018). In addition, there are various remote sensing instruments which are configured to acquire data about the surface temperatures at medium spatial resolutions. The datasets derived from these instruments are beneficial for mapping malaria since a sufficient number of studies have highlighted the correlation between malaria transmission and the ambient surface temperatures (Lunde *et al.*, 2013; Ngarakana-Gwarisa *et al.*, 2014; Blandford *et al.*, 2013; Wardrop *et al.*, 2013). These environmental variables mainly affect disease transmission at local scales, while at a global scale the climatic factors such as rainfall and relative humidity control the malaria transmission patterns – thus contributing greatly to spatial modelling of malaria at that scale (Tanser *et al.*, 2003; Caminade *et al.*, 2014). Remote sensing techniques have widely been used to analyse malaria transmission patterns at various scales, and sometimes there has been some contrasting findings on the correlations between remotely-derived variables with malaria distribution. However, it is very crucial to understand the relationship between remote sensing data and the malaria breeding sites (water/soil) and the questing/resting sites (vegetation/settlements) for accurate malaria mapping, allocation of scarce financial resources and to inform decision-making regarding malaria control interventions. In addition, the understanding of environmental factors associated with malaria transmission will form part of the malaria information management system for enhancing efforts aimed at eliminating malaria by 2020 in South Africa.

In the following section, a comprehensive review on the application of earth observation technology (remote sensing) for malaria studies globally is provided. Firstly, the relationship between malaria vector breeding sites/habitats and reflectance spectroscopy are outlined and discussed. Secondly, a comprehensive description of available satellite sensor portfolio and statistical methods used for mapping malaria is provided. Thirdly, challenges relating to current approaches to mapping malaria through earth observation technology are outlined. Lastly, this chapter seeks to highlight the opportunities presented by earth observation technology for accurate and efficient malaria mapping at various scales.



## 2.2 Malaria vector habitats

All *Anopheline* species depend on the availability of water or water bodies for growth and proliferation. Although water bodies provide suitable habitats for *Anopheline* larval growth, the quantity and the quality of water bodies play a significant role in the malaria transmission rates (Sharma, 2014). By convention, greater risk of malaria is predominantly associated with the frequency, abundance and duration of water bodies. The households located in close proximity to *Anopheline* larval sites are often at greater risk of disease transmission than those located far from the water bodies, especially that the *Anopheline* mosquitoes have limited dispersal range – within 5 km (Kauffman and Briegel, 2004). However, the presence of small water bodies such as natural water-logged areas, burrow pits, open drains, river and dam fringes, temporary puddles, cattle hoofprints, cultivated swamps has been identified as one of the productive breeding sites for *An.gambiae*, *An.fenestus*, *An. bwambiae*, *An. quadriannulatus* and other *An.* species complex (Ndenga *et al.*, 2011; Foley *et al.*, 2003). *An. arabiensis* and *An. fenestus* species for example, prefer to breed in water bodies that are sunlit. This preferred bathymetric characteristic of water bodies reduces the incubation period of *Anopheline* eggs (Impoinvil *et al.*, 2009) because such small water bodies heat quicker than large bodies of water such as the dam (Gibson *et al.*, 2002). Moreover, water quantity is not the only main factor controlling the breeding behaviour of *Anopheline* mosquito. Water quality has also been cited as one of the important hydrological parameters having effect on the *Anopheles* breeding. Most of the species in the *Anopheles* complex breed in clear and stationery water bodies (Dejenie *et al.*, 2011). However, some studies have shown that some species *Anopheline* mosquito have high degree of plasticity which can extend to the habitats previously considered unfavourable. For example, in a study conducted by Gunathilaka *et al.* (2013) it was found that *An.culicifacies* is also adapted to breed in polluted waste water. On the other hand, Sattler *et al.* (2005) have found that *An. gambiae* larvae tend to exist in highly organically contaminated habitats. Larval densities of some *Anopheles* complex (e.g. *An. culicifacies*, *An. fluviatilis*) were found to be associated with biophysical and biochemical parameters such as total hardness, sulphate and chloride, nitrate and calcium in Bahagard and Rudan districts of Iran (Soleimani-Ahmadi *et al.*, 2013).

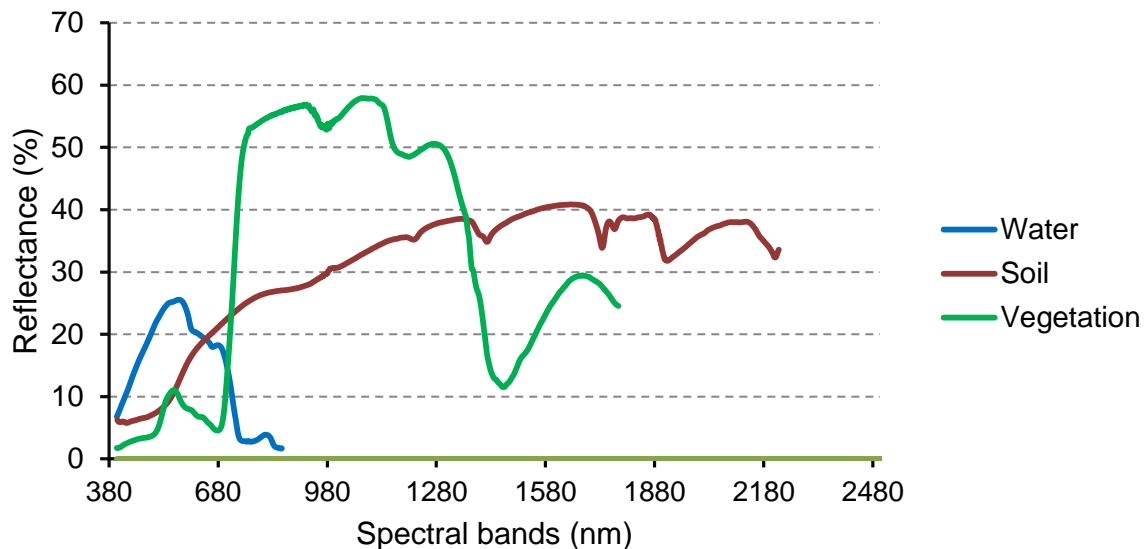
The presence of green vegetation has been suggested as one of the potential indicator of malaria vector habitats. Both aquatic and terrestrial vegetation could provide reliable indicator for mapping of malaria. A study by Gimnig *et al.* (2001) has shown that *An.arabiensis* was associated with temporary habitats containing algae but with no aquatic vegetation, while *An.fenestus* was found in habitats comprising of both algae and aquatic vegetation in Asembo

Bay, western Kenya. The emergent vegetation was found to be the best predictor of *An.gambiae* larval abundance in puddles, tire tracks, ponds and swamps in villages found in three districts of the Kenyan coast (Mwangangi *et al.*, 2007). The presence of green vegetation is also associated with malaria transmission due to its potentially high suitability to *Anopheles* mosquito habitation. The early work of vegetation-mosquito relationship was carried by Service (1971). In this study, it was found that unfed mosquitoes tend to rest in a shelter of dense vegetation, with both male and female species exhibiting distinct preference to vegetation as resting sites. In Costa Rica, Burkett-Cadena *et al.* (2013) have found that *An.albimanus* and other mosquito species were found resting predominantly on understory vegetation than any other resting environments. Similarly, Dewald *et al.* (2016) highlighted that *Anopheles* species require vegetation as a shield against possible desiccation by sunlight in tropical environments. This further emphasizes the importance of vegetation in mosquito growth, pathogen incubation, malaria transmission, prediction and mapping.

### 2.3 Reflectance properties of water

Water is, by far, one of the major regulators of malaria transmission at local, regional and global scales by providing medium through which *Anopheles* mosquito reproduce (Oyewole *et al.*, 2009; Keiser *et al.*, 2005). The availability of water bodies affects malaria vector survival, which includes the survival of *Anopheles* larvae. By highlighting the reflectance properties of water from remote sensing instruments, the contribution of remote sensing in epidemiological studies is thus realized. Pure water exhibits high reflectance in the visible range of the electromagnetic spectrum (380-700nm) and is almost entirely absorbed in the near to far-infrared regions (800 - 2500 nm) (Figure 2.1). However, in instances where water is contaminated by suspended material, chemical compounds, coloured dissolved organic material (CDOM) or nutrients (such as nitrogen, phosphorus, and potassium), the spectral reflectance significantly changes form due to the attenuation of incoming radiation by one or a combinations of these materials. For example, Kong *et al.* (2015) have observed that water bodies with high suspended sediment concentrations tend to exhibit rising spectral response in the visible-near infrared spectrum (400-900 nm) than the water bodies lower suspended sediment concentrations. The variability of sediment concentrations is crucial for the preference of breeding sites for certain mosquito species. More recently, *Anopheles culicifacies* has shown adaptations for breeding in waste water with dissolved organic matter (Gunathilaka *et al.*, 2013; Kelly and Lezaun, 2013). The physio-chemical parameters such as

dissolved nitrate and sulphate concentrations are known to affect the development and survival of *Anopheles* larvae (Mutero *et al.*, 2004; Tchigossou *et al.*, 2017).



**Figure 2.1:** The average spectral reflectance of the three environmental features that affect malaria transmission (water, soil and green vegetation).

Eutrophication of water bodies increases the potential of malaria transmission by providing certain organic nutrients and high likelihood of larval survival. For example, a study by Noori *et al.* (2015) indicated that survival *Culex* mosquito larvae were strongly correlated to the concentrations of  $PO_4$  or  $NO_3$  in polluted water bodies. In eutrophic to hypertrophic water bodies, the spectral reflectance in the visible to near-infrared (380 – 800 nm) mimics that of healthy vegetation due to the chlorophyll concentrations of algae and cyanobacteria (Zhenga and DiGiacomo, 2017; Vos *et al.*, 2003).

#### 2.4 Reflectance properties of soil

Soil plays a major role in the malaria-water-vegetation nexus. Firstly, the frequency, depth and extent of malaria vector breeding sites depend, to a large extent, on the type of soil substrate available in the malaria endemic zones (Machault *et al.*, 2009; Bomblie *et al.*, 2009; Asare *et al.*, 2016; Mentosi *et al.*, 2012). Secondly, soil type and soil bulk density affect the soil water content (moisture) as a function of soil minerals such as aluminium, magnesium and calcium (Gong *et al.*, 2003). Finally, the physical and chemical composition of soil and climate determine the vegetation composition of a landscape (Beukes and Ellis, 2003; Eni *et al.*,



2012). Understanding the spectral reflectance of soil substrate under different environmental/climatic conditions could aid in mapping of malaria especially when the soil parameters are included as malaria predictor variables.

Unlike water, the spectral property of soil extends to the shortwave infrared region of the electromagnetic spectrum (350-2200 nm), depending on whether the soil spectra collected represent moist or dry soil (Figure 2.1). Typically, all soils types have similar spectral curve, with lower reflectance in the visible range (400-650 nm) and higher reflectance in the infrared region, with specific absorption features at 1400, 1900 and 2200 nm (Curcio *et al.*, 2013; Viscarra Rossel *et al.*, 2016). The absorption bands at wavelengths smaller than 1000 nm may be due to iron oxides and chromophores, while the subsequent absorption bands at longer wavelengths near 1400 and 1900 nm result from hydroxyl bonds and water. Additionally, at 2200 nm, the clay minerals sudden dip off the spectral reflectance of the soil. Interestingly, the soil organic matter absorbs and reflects at various wavelengths across the VIS-NIR spectrum (Viscarra Rossel *et al.*, 2016).

The structural and chemical composition of the soil has a direct effect on the formation of the natural malaria breeding sites. Some species of *Anopheles* (e.g. *An. arabiensis*, *An. fenestus*, *An. gambiae*) prefer to breed in small, well-lit water bodies with ambient temperatures, such as the puddles (Ndenga *et al.*, 2011; Ye-Ebiyo *et al.*, 2003) and cattle hoofprints (Paul *et al.*, 2018; Imbahale *et al.*, 2011; Ndenga *et al.*, 2011). The longevity and the structural fabric of these breeding sites depends also on the structure of the soil, as puddles and hoof prints will sustain longer in clayey soil than in sandy soil. This in turn affects water retention capacity and the hydraulic state of the soils on which breeding sites are formed (Romero *et al.*, 2011). In clayey soils, cattle hoofprints could take up to 247 days before they disintegrate (Waudby and Petit, 2015). From a spectroscopy point of view, there is a clear distinction between sandy soil and clay soils, particularly at the spectral regions 500-2100 nm of pre-processed reflectance curves (.e.g. through continuum removal) (Wright *et al.*, 2016). The pedological parameters including soil type, organic matter and water retention capacity have a strong influence on the type and the distribution of vegetation across the landscape. Although the reflectance property of soil varies less, various factors can influence the shape of reflectance curve. Factors such as texture, surface roughness, soil organic matter, spectral mixture, atmospheric conditions and soil moisture variability (Ben-Dor *et al.*, 2003; Wu *et al.*, 2009; Gomez *et al.*, 2008; Nawar *et al.*, 2016; Ben-Dor *et al.*, 2009). These factors should usually be taken into consideration when modelling the spatial variability of soil in relation to malaria vector breeding sites.



## 2.5 Reflectance properties of vegetation

Vegetation has been cited as one of the major parameters associated with malaria transmission patterns in tropical, arid and semi-arid environments (Ricotta *et al.*, 2014; Shiff *et al.*, 2013; Rubio-Palis *et al.*, 2013; Rahman *et al.*, 2011). For malaria vector, vegetation serves as both attractant and a repellent, thus making the contribution of vegetation to malaria transmission complex, particularly at localized scales. Some of the well-known species that repel malaria vector include, among others, ethnobotanical plants such as *Lantana camara*, *Omicum americanum*, *Tagetes minuta*, *Hyptis suaveolens*, *Lippia javanica*, *L. uckambensis*, *A. indica*, and *O. kilimandscharicum* (Seyoum *et al.*, 2002; Mabogo, 1990). On the other hand, a study by Sant'ana *et al.* (2014) has shown that certain species of *Aedes* mosquitoes were attracted to infusions of *Panicum maximum* grass, and areas that comprised of this grass species tend to yield higher oviposition by *Aedes* species than the areas without vegetation. De *et al.* (2017) concluded that the *An. arabiensis* and *An. coluzzi* maintained certain levels of hierarchical preference of grass volatiles for oviposition (Asmare *et al.*, 2017), and such environmental cue could be used to locate suitable breeding sites of each *Anopheline* species. In higher plants, *Psidium guajava* (guava) was found to contain sugar odours that attracted certain species of *An. gambiae* in Mali (Müller *et al.*, 2010).

Measurements of major light absorbing constituents in fruit, flowers, and leaves indicate the significance of these biochemical components as fundamental regulators of vegetation reflectance properties at various levels (leaf and canopy) (Adjorlolo *et al.*, 2012). Generally, the reflectance of vigorously green vegetation in the VIS-SWIR region (380-2500 nm) is very different from that of water and soil in many ways: Firstly, in the visible region (380-700 nm), healthy green vegetation has two main absorption bands in the blue (430 nm) and red (650 nm) regions. The apparent spectral dips occur mainly the absorption of these photons by the chlorophylls for photosynthetic activities (Terashima *et al.*, 2009). The green band is usually higher in reflectance than both the blue and red because green radiation is less absorbed by plants. The reflectance usually peaks in the NIR region, and this is reflectance is due to the scattering by the spongy mesophyll and intercellular spaces (Figure 2.1) (Brodersen and Vogelmann, 2007). In the short-wave infrared (SWIR) region much of the absorption occurs in the water bands (1400, 1900 and 2400 nm), mainly due to vegetation moisture content at leaf or canopy levels (Goward, 1985; Wang and Qu, 2007). By observing the behaviour of spectral response curve of vegetation it is thus possible to measure vegetation stress (Orych

*et al.*, 2013; Sun *et al.*, 2014), vegetation health (Peng *et al.*, 2013), vegetation community type (Liu *et al.*, 2016), tree species (Jiménez and Dáz-Delgado, 2015), biomass (Thapa *et al.*, 2014), and irrigation status of cultivated area (Conrad *et al.*, 2007).

Many factors affect the spectral response of vegetation particularly in the VIS-SWIR spectrum (350-2500 nm). The light interception by vegetation canopy depends upon the morphology of leaves (leaf area), the leaf angle, and sun-viewing geometry (Jacob *et al.*, 2007; Sellers, 1985). In addition to the structural morphology of leave, canopy and optical properties of vegetation, the soil background often contributes to the spectral reflectance of the target especially when using decadal (10 m, 20 m etc.) sensors (Munyati *et al.*, 2013). Huete (1988) has alluded that soil background exerts significant effect on the reflectance of vegetation, and this could ultimately result in incorrect derivation of vegetation indices in the VIS-NIR region. On the other hand, variations of the soil moisture alter the spectral reflectance across the water absorption bands of the VIS-SWIR spectrum.

## 2.6 Mapping malaria: remote sensing perspective

One of the early works relating to application of remote sensing technology was reported in 1973 by National Aeronautical and Space Administration where the scientists, in collaboration with the team from New Orleans Mosquito Control District, investigated the potential application of aerial photography for mapping vegetation communities associated with *Aedes sollicitans* habitat (Hay *et al.*, 1997). Since then there were studies that utilized a 2.5 m spatial resolution airborne Multi-Spectral Scanner (MSS) which yielded higher classification accuracies than the colour-infrared aerial photography. In addition to its high spatial resolution (2.5 m), the initial MSS comprised of many spectral bands in the 300-1300 nm (Hay *et al.*, 1997) that improved characteristics of *Aedes* habitats. Such improvements were reported by Barnes and Cibula (1979). Consequently, the improvement in the radiometric, temporal, spatial and spectral resolution enabled mapping of, not only vegetation assemblages relating to malaria transmission, but also quantification of other environmental factors such as water bodies, surface temperatures and soil characteristics associated with breeding sites. Since 1972, there was an increased number of earth observation (EO) satellites launched, with differing spectral and spatio-temporal configuration. Parallel to the launch of many EO satellites was the increase in the adoption of EO uptake for epidemiological studies, particularly mapping of infectious diseases with high morbidity. In addition, many other EO instruments were also applied in agriculture, soil science, disaster management, forestry,

spatial planning and many other areas (Table 2.1), which directly or indirectly affected malaria transmission.

Both precipitation and daily/diurnal temperatures have shown significant correlations with the malaria transmission and such correlations have been reported by many researchers over the years. The use of TRMM, which is a meteorological satellite, has shown tremendous success in malaria mapping for many years. Research conducted by Alegana *et al* (2013) has successfully utilized monthly precipitation data derived from TRMM to estimate malaria incidence in Namibia. From this research, it was found that precipitation (0.02, 0.01-0.03,  $p = 0.02$ ), temperature suitability index (7.57, 5.34-9.96,  $p = 0.001$ ) and the enhanced vegetation index (6.55, 4.25-8.87,  $p = 0.001$ ) were significant predictors of crude malaria incidence rate. In Southern Africa, many studies have been conducted to evaluate the usefulness of the climate data derived from remote sensing instruments such as TRMM. Komen (2017) and Chuang *et al.* (2017) have recently indicated that TRMM variables were significant predictors of during the malaria onset and can thus be used as input for malaria elimination control efforts in South African and Swaziland (now called *eSwatini*) respectively. The complementation of very coarse spatial resolution climate dataset with medium to high resolution satellite datasets has contributed significantly to the understanding of environmental factors associated with local malaria transmission. For example, a recent study by Ferrao *et al.* (2018) has used climate data with vegetation proxy (NDVI) and land use/land cover data derived from Landsat, to map the population at risk of malaria in Chimoio, Mozambique.

The advent of very high spatial resolution satellite data such as the SPOT (*Satellite Pour l'Observation de la Terre*) has opened new opportunities for mapping of malaria at much higher accuracies than its Landsat predecessor. For example, Machault *et al.* (2012) have mapped the *An. gambiae* densities in Dakar using environmental variables derived from SPOT 5 sensor. In this study, it was concluded that SPOT dataset provided the first example of high resolution EO products which can be integrated in operational early warning system for malaria vector control. More recently, Minale and Alemu (2018) have tested the use of both SPOT and SRTM to map malaria risk in Bahir Dar City, Ethiopia. The findings from this study showed that only 2% of the land area under the city administration was risk-free with large proportion of city area (65%) under high risk of malaria infection. The potential of IKONOS dataset with the spatial resolution of 0.82 cm spatial resolution has been explored for intra-urban malaria risk mapping in Dar es Salaam. In this study, Kabaria *et al.* (2016) have found that the percentage of dense/riverine vegetation, proximity to water bodies and built-up areas



derived from IKONOS data were significant predictors of *P. falciparum* prevalence rate in children of ages between 2 and 10 years old.



**Table 2.1:** Some of the major globally available EO satellites to date. Satellites are arranged in alphabetical order and not according to importance.

Satellite name	Launched	Spatial resolution	Spectral resolution	Revisit time	Applications (References)
Advanced Land Observation Satellite (ALOS)	January 2006	2.5 m	420-890 nm	46 days	Crop estimations ( <a href="#">Zhang et al., 2009</a> ), wetland mapping and change analysis ( <a href="#">Rebelo et al., 2009</a> ), forest mapping ( <a href="#">Thapa et al., 2014</a> )
Advanced Very High Resolution Radiometer (AVHRR)	October 1978	1.1 km	580-12 500 nm	1 day	Soil mapping ( <a href="#">Dobos et al., 2001</a> ), malaria mapping ( <a href="#">Rahman et al., 2010</a> ), water applications ( <a href="#">Dietz et al., 2017</a> )
China-Brazil Earth Resource Satellite (CBERS-3/4)	December 2014	20 m	420-890 nm	26 days	Impact assessment ( <a href="#">Yang et al., 2012</a> ), water quality assessment ( <a href="#">Zhang et al., 2010</a> ), soil mapping ( <a href="#">Pereira et al., 2017</a> )
Earth Observing-1 Mission (EO-1)	November 2000	30 m	400-2500 nm	16 days	Disease mapping ( <a href="#">Hamm et al., 2015</a> ), water quality/quantity assessment ( <a href="#">Gong et al., 2014</a> <a href="#">Mohamed, 2015</a> ), soil mapping ( <a href="#">Bannari et al., 2016</a> ).
IKONOS	September 1999	1 m	450-900 nm	3 days	Vegetation mapping ( <a href="#">Wang et al., 2011</a> ), malaria risk mapping ( <a href="#">Krefis et al., 2011</a> ), water mapping ( <a href="#">Mishra et al., 2004</a> )
Landsat 1(MSS), 5-7 (TM,ETM+) 8 (OLI)	July 1972 (MSS); March 1984 (TM); April 1999 (ETM+); February 2013 (OLI)	30 m	435-12 510 nm	16 days	Soil classification ( <a href="#">Nikolaos, 1988</a> ), land use mapping ( <a href="#">Hu et al., 2016</a> ), malaria mapping ( <a href="#">Adeola et al., 2017</a> ; <a href="#">Malahlela et al., 2018</a> ), soil mapping ( <a href="#">Scudiero et al., 2015</a> ).
Meteosat	November 1995	2.5 km	500-12 500 nm	15 minutes	Soil moisture mapping ( <a href="#">Wagner et al., 2007</a> ; <a href="#">Verstraeten et al., 2006</a> ), malaria risk modelling



Moderate-Resolution Imaging Spectroradiometer (MODIS)	December 1999	250 m	405-14 385 nm	1 day	( <a href="#">Omumbo et al., 2005</a> ), vegetation monitoring ( <a href="#">Ghilain et al., 2014</a> ) Fire mapping ( <a href="#">Giglio et al., 2003</a> ), irrigation monitoring ( <a href="#">Conrad et al., 2007</a> ), human settlement mapping ( <a href="#">Schneider et al., 2010</a> )
Project for On-board Autonomy (PROBA)	October 2001	8 m	400-1050 nm	< 7 days	Crop mapping ( <a href="#">Lambert et al., 2016</a> ), hydrological mapping ( <a href="#">Wirion et al., 2017</a> )
Quickbird	October 2001	0.65 m	450-900 nm	1-3.5 days	Malaria mapping ( <a href="#">Jacob et al., 2007</a> ), soil mapping ( <a href="#">Sidike et al., 2014</a> ), disaster mapping ( <a href="#">Kerle, 2010</a> ), weed management ( <a href="#">Lopez-Granados, 2011</a> )
Rapid Eye	August 2008	6.5 m	440-850 nm	5.5 days	Soil chemistry modelling ( <a href="#">Blasch et al., 2015</a> ), vegetation mapping ( <a href="#">Ramoelo et al., 2015</a> ), aquatic mapping ( <a href="#">Fritz et al., 2017</a> )
Resource-SAT	October 2003	24 m (LISS-III); 56 m (AWiFS)	0.52-1 700 nm	24 days	Disaster management ( <a href="#">Bahuguna et al., 2008</a> ), snow mapping ( <a href="#">Kulkarni et al., 2006</a> ), groundwater prospecting ( <a href="#">Vijith, 2007</a> )
Satellites Pour l'Observation de la Terre (SPOT) 1-7	February 1986	5 m (SPOT 1-5); 1.5 m (SPOT 6,7)	455-1750 nm	1-2 days	Malaria mapping ( <a href="#">Kabaria et al., 2016</a> ), soil mapping ( <a href="#">Sumfleth and Duttmann, 2008</a> ), vegetation estimation ( <a href="#">Sha et al., 2016</a> )
Sentinel-1, 2A/B	June 2015	10 m	443-2190 nm	5 days	Soil mapping ( <a href="#">van der Werff and van der Meer, 2015</a> ), forest mapping ( <a href="#">Sothe et al., 2017</a> ), water body mapping ( <a href="#">Du et al., 2016</a> )
Shuttle Radar Topography Mission (SRTM)	February 2000	30 m	7.5-3.5 cm	11 days	Disease mapping ( <a href="#">Mosomtai et al., 2018</a> ), soil mapping ( <a href="#">Forkuor et al., 2017</a> ), vegetation characterization ( <a href="#">Kelindorfer et al., 2004</a> )
Tropical Rainfall Measuring Mission (TRMM)	November 1997	4.3 km (PR); 2.2 km	630-12 000 nm (VIRS); 0.3-100 µm (CERES)	46 days (CERES)	Vegetation monitoring ( <a href="#">Zhang et al., 2005</a> ), drought monitoring ( <a href="#">Du et al., 2013</a> ), soil moisture



---

		(VIRS); 4.4 (TMI)			studies ( <a href="#">Bindlish et al., 2003</a> ), malaria incidence mapping ( <a href="#">Alegana et al., 2013</a> )
WorldView-2/3	October 2009	0.46 m	410-1050 nm	1.1 days	Vegetation mapping ( <a href="#">Yan et al., 2018</a> )

PR = precipitation radar; MSS = multispectral scanner; VIRS = visible and infrared scanner; TMI = TRMM microwave imager; CERES = clouds and the earth's radiant energy system; LISS = linear imaging self-scanning sensor; AWiFS = advanced wide field sensor; TM = thematic mapper; ETM+ = enhanced thematic mapper plus; OLI = operational land imager.

---

Although many studies have reported success in mapping malaria with EO datasets, there are still challenges that can be grouped into two main categories: (i) data challenges, and (ii) challenges relating to mapping methods.

## 2.7 Malaria mapping: data challenges

Malaria is transmitted to humans through the bite of the infected female mosquito. There are various pathogens that are transmitted by different mosquitoes which belong to the family *Anophelinae*. The well-known deadly malaria pathogen is the *Plasmodium falciparum*, which is the protozoa parasite for humans that is largely distributed in the warm, moist tropics of the world (Gething *et al.*, 2011). Malaria season in tropical environments usually coincides with the periods of heavy rainfall and high temperature, which eventually reduces the visibility of bottom-of-atmosphere through optical remote sensing (350-2500 nm) (Wang *et al.*, 1999). The presence of clouds and cloud shadows impedes the application of optical remote sensing. For example, cloud cover and cloud shadowing affect the spectral response of potential malaria vector habitat, thereby over-estimating or underestimating the distribution of such habitats (Zhang *et al.*, 2015). This makes the application of optical sensing limited to mainly pre and post analysis of malaria mapping and seldom during malaria season.

Although a number of remote sensing datasets (ranging from multispectral to hyperspectral) are available, processing of such datasets requires specialized and computationally efficient methods. Unfortunately, malaria is common mostly in environments where there is limited expertise in terms of image processing and such areas often have poor infrastructure to handle large volumes of spatial datasets. The retrieval of biophysical and biochemical characteristics related to malaria breeding and resting sites requires the collection of spectral signatures from different images at different temporal and spatial resolutions. The very high resolution datasets often cover smaller areas and the extension to wide-area image acquisition often incurs exorbitant costs (Fisher *et al.*, 2017). On the other hand, there should be careful selection of spectral bands that are sensitive to the objects of interest – failure to which results in redundancy of data even for the high resolution imagery.



## 2.8 Methodological challenges

### 2.8.1 Classification

Many researchers have applied the supervised classification techniques on satellite imagery to characterize malaria vectors' suitable habitats. The land use and land cover map, for example, is the most commonly remote sensing product that is often used in malaria studies. A study by [Mohan and Naumova \(2014\)](#) has shown that an increase in malaria cases was proportional to LULC change between 2000 and 2003. [Dlamini et al. \(2015\)](#) derived LULC from RapidEye imagery to assess its relationship with the malaria vector breeding sites in eSwatini. However, habitat delineation through image classification techniques is faced with major challenges associated with satellite data. The major challenge relates to definition of adequate hierarchical levels for mapping land cover units discernable by selected remote sensing data. In addition, selecting adequate and 'pure' representative training sites may also be a limiting factor in image classification ([Lu and Weng, 2007](#)). Additionally, post-classification processing of satellite imagery requires expert knowledge particularly considering the 'salt-and-pepper' effect of per-pixel classification ([Lu and Wang, 2004](#)).

### 2.8.2 Regression

Regression is the widely used method for relating the environmental/climatic variables with malaria vector habitats. Most studies have used logistic regression model (a form of generalized linear model) for linking the habitat distribution of malaria vector with environmental and remote sensing indices. A study by [Mushinzimana et al. \(2006\)](#) demonstrated the derivation of the link landscape determinants and *Anopheles* mosquito larval habitats in Kenya highlands by means of stepwise logistic regression. In this study, various LULC classes have shown correlation with aquatic habitats of *Anopheles* mosquitoes. In logistic regression, the binary outcome (presence or absence) of habitat suitability is modelled as a function of the coefficients of predictor variables (in this case remote sensing variables) using a logit link function. Although the logistic regression model has been used widely in habitat suitability mapping of malaria vectors, there are widely recognised challenges associated with this model. One of the challenges is the number of variables, the algorithm used (backward selection, forward selection or both) and the order in which variables are entered in the model (or removal) can all affect the selected model. This may in turn affect the model robustness especially when there is a high degree of multicollinearity amongst predictor variables which goes untreated ([Grafen and Hails, 2002](#)). Another challenge arises when inference is based on the global model, where tests of individual parameters (variable importance) are conducted using the null hypothesis testing – a hypothesis testing which has

since received much criticism in recent times (Anderson *et al.*, 2000). Generally, a model selection can be a tedious exercise particularly when multiple performing analytical pre-processing of input parameters, which is often not reported in modelling.

## 2.9 Opportunities

Mapping malaria is based on the relationship of *Anopheles* breeding and resting sites with the geospatial datasets. The assumption made about the potential mosquito habitats is that areas of high habitat suitability have unique spectral characteristics defined by biophysical and biochemical composition compared to the unsuitable habitats. Such uniqueness forms the basis for discriminating areas suitable for malaria at various spatial resolutions. For example, it has been reported that potential habitats of malaria vector can be delineated from satellite datasets by using spectral reflectance of vegetation land cover and irrigated croplands (Jacob *et al.*, 2007; Baeza *et al.*, 2013). Some of the crucial characteristics of vegetation including nutrients (Ramoelo and Cho, 2018), vegetation canopy moisture (Malahlela *et al.*, 2018), plant pigments such as chlorophylls and carotenoids (Gitelson *et al.*, 2003), biomass (Prabhakara *et al.*, 2015), species diversity and composition (Madonsela *et al.*, 2017), and leaf area index (Roosjen *et al.*, 2018) predominantly affect the spectral reflectance vegetated malaria vector habitats. This, therefore, makes it challenging to accurately quantify potential habitats of malaria vector especially because the spectral mixture of such habitats is not only defined by the vegetation biophysical/chemical characteristics, but also the climatic factors and background materials within the vicinity of such habitats. This makes mapping of malaria challenging, particularly considering the soil-water-vegetation nexus nature of malaria vector habitats.

The improvement in radiometric, spectral and spatio-temporal configuration of current and future EO satellites offers new opportunities for improved quantification of *Anopheles* habitats. This thus makes it possible to untangle various biophysical and biochemical characteristics of mosquito breeding (water) and resting (vegetation) sites that was, otherwise, difficult to quantify using traditional remote sensing datasets. For example, the presence of narrow-band such as the red edge in WorldView-2, RapidEye, and Sentinel-2 offers opportunities to characterize variations in physico-chemical constitution of soil-water-vegetation nexus. For example, Kross *et al.* (2015) found that the vegetation indices computed from the red-edge band performed consistently better than the traditional red band indices when estimation vegetation leaf area index (LAI). However, traditional classification methods designed for

conventional datasets are unable to produce consistent and robust results when mapping malaria, due to the high data dimensionality of EO datasets such as the hyperspectral data. The improvements in sensor specifications also demand rapid algorithms and infrastructure geared towards handling and processing high data volumes.

Many studies have shown that the machine learning algorithms such as support vector machine (SVM) and random forest (RF) higher performance for both classification and regression analysis of data. For example, [Tian \*et al.\* \(2016\)](#) have utilized both SVM and RF on multi-sensor datasets for classifying wetlands, which are primarily associated with malaria vector breeding sites in tropical regions of the world ([Sousa \*et al.\*, 2009](#)). On the other hand, [Kapwata and Gebreslasie \(2016\)](#) have utilized RF for selecting variables for spatially modelling malaria transmission in Mpumalanga province of South Africa. The availability of machine learning algorithms and high resolution dataset offers opportunities for in-depth malaria studies, for understanding the biophysical and biochemical characteristics of malaria habitats and to predict malaria incidence rates in endemic areas. Additionally, by fusing conventional classifiers and datasets, it thus becomes possible to map malaria incidences at much higher accuracies. A recent study by [Li \*et al.\* \(2016\)](#) has fused conventional image classification method with knowledge based model to map malaria hazard at landscape level. In this model, various matrices of land use and land cover (LULC) were derived from SPOT 5 satellite images.

## **2.10 Further research requirements in remote sensing for malaria mapping**

Although there has been success in quantifying both vegetated and moist habitats of malaria vector through remote sensing, there is a lack of high resolution characterization of such habitats which takes into account the micro-climate of both breeding and resting sites. Mapping of micro-habitats such as cattle hoofprints and puddles has not been done and as such this is invariably a research gap that exists in malariometric studies. In the past, one of the limitations to the mapping the micro-habitats of malaria vector (cattle hoofprint and puddles) was the spectral and spatial resolution of the commonly used EO datasets. The newly-launched high resolution satellites such as WorldView-2/3 and Sentinel 2 offer an opportunity to map mosquito breeding and resting sites efficiently and accurately. In ecology, the control of malaria by targeting the elimination of mosquitoes is not necessary because at any time once the vector that incubates *P. falciparum* has been eliminated, a more potent one emerges ([Killen \*et al.\*, 2013](#)). This is particularly exacerbated by climate change which

increases the risk of malaria re-emergence in endemic and epidemic areas (Ivanescu *et al.*, 2016). The re-emergence of malaria has been reported elsewhere (Sharma, 1996) due to shortage of the dichlorodiphenyltrichloroethane (DDT) commonly used for malaria vector control. Although the DDTs play a crucial role for rapid control against malaria in many countries, the potential hazards associated with their continued use were first reported in 1944 (Davis, 2014). The apparent hazard posed by the use of DDT mainly due to the high levels of dosage, is the ability to alter the functioning of nervous system in humans and domestic animals. This could result in dizziness, convulsions, tremor and instability as a result of tissue poisoning by DDT (Katole *et al.*, 2013). Due to the toxicity and cost associated with the use of DDTs, many people in communal areas have adopted the use of ethnobotanical plants for malaria control. In South Africa, one of the most common of such species is *Lippia javanica* (Lemon bush) which is widely used for its aromatic effect that serves as a repellent for mosquitoes. On the other hand, quantification of plant species used as ethnobotanical plants for controlling *Anopheles* mosquito is necessary for comprehensive approach to malaria mapping and control.

## 2.11 Conclusions

Mapping the potential habitats for malaria vectors is one of the challenging steps for malaria control strategies. Since the realization of the contribution of vegetation, soil and water to malaria incidence rates, attempts have been made to relate such habitats with remote sensing approaches. However, from this literature it becomes evident that most of the studies that utilized remote sensing relied on the indication given by vegetation greenness indices (NDVI, for example) while neglected some vegetation parameters which are very closely linked to mosquito's resting/questing behaviour. The high resolution spatial information on the distribution of mosquito micro-habitats such as cattle hoofprints and puddles (natural or man-made), which are important for malaria transmission is still missing, to a greater extent. On the other hand, characterizing vegetation parameters that largely dictate the survival of adult mosquitoes (e.g. LAI) in relation to malaria has not been adequately done, and as such more work is required to assess the potential of high resolution satellite dataset for estimating LAI in areas endemic to malaria. The following chapter will discuss remote sensing approaches for mapping the cattle hoofprints which also serve as some of the habitats for *Anopheles* mosquitoes in subtropical areas.

## 2.12 References

- Adeola A.M., Botai J.O., Olwoch J.M., Rautenbach C.J., Adisa O.M., Taiwo O.J., Kalumba A.M. (2016). Environmental factors and population at risk of malaria in Nkomazi municipality, South Africa. *Tropical Medicine and International Health*, 21(5): 675 – 686.
- Adeola A.M., Botai, J.O., Rautenbach H., Adisa O.M., Ncongwane K.P., Botai C.M., Adebayo-Ojo T.C. (2017). Climatic Variables and Malaria Morbidity in Mutale Local Municipality, South Africa: A 19-Year Data Analysis. *International Journal of Environmental Research and Public Health*, 14(11):1360
- Adjorlolo C., Mutanga O., Cho M.A., Ismail R. (2012). Challenges and opportunities in the use of remote sensing for C3 and C4 grass species discrimination and mapping. *African Journal of Range and Forage Science*, 29(2): 47 – 61.
- Alegana V.A., Atkinson P.M., Wright J.A., Kamwi R., Uusiku P., Katokele S., Snow R.W., Noor A.M. (2013). Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial–temporal models. *Spatial and Spatio-temporal Epidemiology*, 7: 25 – 36.
- Alimi T.O., Fuller D.O., Herrera S.V., Arevalo-Herrera M., Quinones M.L., Stoler J.B., Beier J.C. (2016). A multi-criteria decision analysis approach to assessing malaria risk in northern South America. *BMC Public Health*, 16:221. DOI: 10.1186/s12889-016-2902-7.
- Amadi J.A., Olago D.O., Ong’amo G.O., Oriaso S.O., Nanyingi M., Nyamongo I.K., Estambale B.B.A. (2018). Sensitivity of vegetation to climate variability and its implications for malaria risk in Baringo, Kenya. *PLOS One*, 13(7): e0199357. DOI: [10.1371/journal.pone.0199357](https://doi.org/10.1371/journal.pone.0199357).
- Anderson D.R., Burnham K.P., Thompson W.L. (2000). Null hypothesis Testing: Problems, Prevalence and an alternative. *The Journal of Wildlife Management*, 64(4): 912 – 923.
- Asare E.O., Tompkins A.M., Amekudzi L.K., Emert V. (2016). A Breeding Site Model for Regional, Dynamical Malaria Simulations Evaluated Using *In Situ* Temporary Ponds Observations. *Geospatial Health*, 11(1s): DOI: 10.4081/gh.2016.390.
- Asmare Y., Hill S.R., Hopkins R.J., Tekie H., Ignell R. (2017). The role of grass volatiles on oviposition site selection by *Anopheles arabiensis* and *Anopheles coluzzii*. *Malaria Journal*, 16:65. DOI: 10.1186/s12936-017-1717-z. (Page 37)
- Baeza A., Bouma M.J., Dhiman R.C, Bakerville E.B., Ceccato P., Yadav R.S., Pascual M. (2013). Long-lasting transition toward sustainable elimination of desert malaria

under irrigation development. *Proceedings of the National Academy of Sciences of the United States of America*, 110(37): 15157 – 15162.

- Bahuguna A., Nayak S., Roy D. (2008). Impact of the tsunami and earthquake of 26th December 2004 on the vital coastal ecosystems of the Andaman and Nicobar Islands assessed using RESOURCESAT AWiFS data. *International Journal of Applied Earth Observation and Geoinformation*, 10(2): 229 – 237.
- Bannari A., Guedon A.M., El-Ghmani A. (2016). Mapping Slight and Moderate Saline Soils in Irrigated Agricultural Land Using Advanced Land Imager Sensor (EO-1) Data and Semi-Empirical Models. *Communications in the Soil Science and Plant Analysis*, 47(16): 1883 – 1906.
- Barnes C.M., Cibula W.G. (1979). Some implications of remote sensing technology in insect control programs including mosquitoes. *Mosquito News*, 39: 271 – 282.
- Ben-Dor E., Chabrilat S., Dematte J. A. M., Taylor G.R., Hill J., Whiting M. L., Sommer S. (2009). Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment*, 113: 538–555.
- Ben-Dor E., Goldshleger N., Benyamini Y., Agassi M., Blumberg D.G. (2003). The Spectral Reflectance Properties of Soil Structural Crusts in the 1.2- to 2.5- $\mu$ m Spectral Region. *Soil Science Society of American Journal*, 67: 289 – 299.
- Beukes P.C., Ellis F. (2003). Soil and vegetation changes across a Succulent Karoo grazing gradient. *African Journal of Range and Forage Science*, 20(1): 11 – 19.
- Bindlish R., Jackson T.J., Wood E., Gao H., Starks P., Bosch D., Lakshmi V. (2003). Soil moisture estimates from TRMM Microwave Imager observations over the Southern United States. *Remote Sensing of Environment*, 85(4): 507 – 515.
- Blandford J.I., Blandford S., Crane R.G., Mann M.E., Paaijmans K.P., Schreiber K.V., Thomas M.B. (2013). Implications of temperature variation for malaria parasite development across Africa. *Scientific Reports*, 3(1300): DOI:10.1038/srep01300.
- Blasch G., Spengler D., Itzerott S., Wessolek G. (2015). Organic Matter Modelling at the Landscape Scale Based on Multitemporal Soil Pattern Analysis Using RapidEye Data. *Remote Sensing*, 7(9): 11125 – 11150.
- Bobos E., Montanarella L., Negre T., Micheli E. (2001). A regional scale soil mapping approach using integrated AVHRR and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, 3(1): 30-42.
- Bomblies A., Duchemin J.B., Eltahir E.A.B. (2009). A mechanistic approach for accurate simulation of village scale malaria transmission. *Malaria Journal*, 8(223): DOI: 10.1186/1475-2875-8-223.





- Brodersen C.R., Vogelmann T.C. (2007). Do epidermal lens cells facilitate the absorptance of diffuse light? *American Journal of Botany*, 94:1061-1066
- Burkett-Cadena N., Graham S.P., Giovanetto L.A. (2013). Resting environments of some Costa Rican mosquitoes. *Journal of Vector Ecology*, 38(1): 12 – 19.
- Caminade C., Kovats S., Rocklov J., Tompkins A.M., Morse A.P., Colon-Gonzalez F.J., Stenlund H., Martens P., Lloyd S.J. (2014). Impact of climate change on global malaria distribution. *PNAS*, 111(9): 3286 – 3291.
- Chuang T., Soble A., Ntshalintshali N., Mkhonta N., Seyama E., Mthethwa S., Pindolia D., Kunene S. (2017). Assessment of climate-driven variations in malaria incidence in Swaziland: toward malaria elimination. *Malaria Journal*, 16(232). DOI: 10.1186/s12936-017-1874-0.
- Conrad C., Dech S.W., Hafeez M., Lamers J., Martius C., Strunz G. (2007). Mapping and assessing water use in a Central Asian irrigation system by utilizing MODIS remote sensing products. *Irrigation and Drainage Systems*, 21(3-4): 197 – 218.
- Curcio D., Ciraolo G., Asaro F.D., Minacapilli M. (2013). Prediction of soil texture distributions using VNIR-SWIR reflectance spectroscopy. *Procedia Environmental Sciences*, 19: 494 – 503.
- Dalrymple U., Mappin B., Gething P.W. (2015). Malaria mapping: understanding the global endemicity of *falciparum* and *vivax* malaria. *BMC Medicine*, 13(140): <https://doi.org/10.1186/s12916-015-0372-x>
- Davis F.R. (2014). *Banned: A history of pesticides and the science of toxicology*. Yale University Press. Pp 26. ISBN 978-0300205176. Retrieved 25 October 2018.
- Dejenie T., Yohannes M., Assmelash T. (2011). Characterization of Mosquito Breeding Sites in and in the Vicinity of Tigray Microdams. *Ethiopian Journal of Health Science*, 21(1): 57 – 66.
- Dewald J.R., Fuller D.O., Muller G.C., Beier J.C. (2016). A novel method for mapping village-scale outdoor resting microhabitats of the primary African malaria vector, *Anopheles gambiae*. *Malaria Journal*, 15(489): DOI: 10.1186/s12936-016-1534-9.
- Dietz A.J., Klein I., Gessner U., Frey C.M., Kuenzer C., Dech S. (2017). Detection of Water Bodies from AVHRR Data—A TIMELINE Thematic Processor. *Remote Sensing*, 9(57): DOI: :10.3390/rs9010057.
- Dlamini S.N., Franke J., 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(1): DOI: 10.4081/gh.2015.302.

- Dobos E., Montanarella L., Negre T., Micheli E. (2001). A regional scale soil mapping approach using integrated AVHRR and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, 1: 30-42.
- Du L., Tian Q., Yu T., Meng Q., Jancso T., Udvardy P., Huang Y. (2013). A comprehensive drought monitoring method integrating MODIS and TRMM data. *International Journal of Applied Earth Observation and Geoinformation*, 23: 245 – 253.
- Eni D.D., Iwara A.I., Offiong R.A. (2012). Analysis of Soil-Vegetation Interrelationships in a South-Southern Secondary Forest of Nigeria. *International Journal of Forestry Research*, 469326: DOI: 10.1155/2012/469326.
- Ferrao J.L., Niquisse S., Mendes J.M., Painho M. (2018). Mapping and Modelling Malaria Risk Areas Using Climate, Socio-Demographic and Clinical Variables in Chimoio, Mozambique. *International Journal of Environmental Research and Public Health*, 15(796): DOI: 10.3390/ijerph15040795.
- Fisher J.R.B., Acosta E.A., Dennedy-Frank P.J., Kroeger T., Boucher T.M. (2017). Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. *Remote Sensing in Ecology and Conservation*, 4(2): 137 – 149.
- Foley D., Torres E., Mueller I., Bryan J.H., Bell D. (2003). Host-dependent *Anopheles flavirostris* larval distribution reinforces the risk of malaria near water. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 97: 283-287.
- Forkuor G., Houkpatin O.K.L., Welp G., Thiel M. (2017). High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLOS ONE*, 12(1): e0170478. DOI: 10.1371/journal.pone.0170478.
- Fritz C., Dornhofer K., Schneider T., Geist J., Oppelt N. (2017). Mapping Submerged Aquatic Vegetation Using RapidEye Satellite Data: The Example of Lake Kummerow (Germany). *Water*, 9(7): 510. DOI: 10.3390/w9070510.
- Gething P.W., Patil A.P., Smith D.L., Guerra C.A., Elyazar R.F., Johnston G.L., Tatem A.J., Hay S.I. (2011). A new world malaria map. *Plasmodium falciparum* endemicity in 2010. *Malaria Journal*, 10(378): <https://doi.org/10.1186/1475-2875-10-378>.
- Ghilain N., De Roo F., Gellens-Meulenberghs F. (2014). Evapotranspiration monitoring with Meteosat Second Generation satellites: improvement opportunities from moderate spatial resolution satellites for vegetation. *International Journal of Remote Sensing*, 35(7): 2654 – 2670.





- Gibson R.N., Robb L., Wennhage H., Burrows M.T. (2002). Ontogenetic changes in depth distribution of juvenile flatfishes in relation to predation risk and temperature on a shallow-water nursery ground. *Marine Ecology Progress Series*, 229: 233 – 244.
- Giglio L., Descloitres J., Justice C.O., Kaufman Y.J. (2003). An Enhanced Contextual Fire Detection Algorithm for MODIS. *Remote Sensing of Environment*, 87(2-3): 273 – 282.
- Gimnig J.E., Ombok M., Kamau L., Hawley W.A. (2001). Characteristics of larval *Anopheles* (Diptera: Culicidae) habitats in Western Kenya. *Journal of Medical Entomology*, 38(2): 282 – 288.
- Gitelson A.A., Gritz Y., Merzlyak M.N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3): 271 – 282.
- Gomez C., Viscarra Rossel R.A., McBratney A.B. (2008). Soil organic carbon prediction by hyperspectral remote sensing and field VIS–NIR spectroscopy: an Australian case study. *Geoderma*, 146: 403–411.
- Gong P., Miao X., Battaglia C., Biging G.S. (2014). Water table level in relation to EO-1 ALI and ETM+ data over a mountainous meadow in California. *Canadian Journal of Remote Sensing*, 30(5). 691 – 696.
- Gong Y., Cao Q., Sun Z. (2003). The effects of soil bulk density, clay content and temperature on soil water content measurement using time-domain reflectometry. *Hydrological Processes*, 17(18): 3601 – 3614.
- Goward S.N. (1985). Shortwave infrared detection of vegetation. *Advances in Space Research*, 5(5): 41 – 50.
- Grafen F.S., Hails R. (2002). *Modern Statistics for the Life Sciences*. Oxford University Press, Oxford.
- Gunathilaka N., Fernando T., Hapugoda M., Wickremasinghe R., Wijeyerathne P., Abeyewickreme W. (2013). *Anopheles culicifacies* breeding in polluted water bodies in Trincomalee District of Sri Lanka. *Malaria Journal*, 12(285): DOI: 10.1186/1475-2875-12-285.
- Hamm N.A.S., Soares Magalhaes R.J., Clements A.C.A. (2015). Earth Observation, Spatial Data Quality, and Neglected Tropical Diseases. *PLOS Neglected Tropical Diseases*, 9(12): e0004164. DOI: 10.1371/journal.pntd.0004164.
- Hay S.I., Packer M.J., Rogers D.J. (1997). The impact of remote sensing on the study of invertebrate intermediate hosts and vectors for disease. *International Journal of Remote Sensing*, 18(14): 2899 – 2930.



- Hay S.I., Snow R.W. (2006). The Malaria Atlas Project: Developing Global Maps of Malaria Risk. *PLOS Medicine*, 3(12): e473. <https://doi.org/10.1371/journal.pmed.0030473>.
- Hu T., Yang J., Li X., Gong P. (2016). Mapping Urban Land Use by Using Landsat Images and Open Social Data. *Remote Sensing*, 8(2): 151. DOI: 10.3390/rs8020151.
- Huete A.R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25: 295 – 309.
- Imbahale S.S., Paaijmans K.P., Mukabana W.R., van Lammeren R., Githeko A.K., Takken W. (2011). A longitudinal study on *Anopheles* mosquito larval abundance in distinct geographical and environmental settings in western Kenya. *Malaria Journal*, 10(91): DOI: 10.1186/1475-2875-10-81.
- Ivanescu L., Bodale I., Florescu S., Roman C., Acatrinei D., Miron L. (2016). Climate Change Is Increasing the Risk of the Reemergence of Malaria in Romania. *BioMed Research International*, ID: 8560519. DOI: <http://dx.doi.org/10.1155/2016/8560519>.
- Jacob B.G., Muturi E.J., Mwangangi J.M., Funes J., Caamano E.X., Muriu S., Shillu J., Githure J., Novak R.J. (2007). Remote and field level quantification of vegetation covariates for malaria mapping in three rice agro-village complexes in Central Kenya. *International Journal of Health Geographics*, 6(21): DOI: 10.1186/1476-072X-6-21.
- Jamison D.T., Breman J.G., Measham A.R., Alleyne G., Claeson M., David B Evans, Jha P, Mills A., Musgrove P. (2006). Disease control priorities in developing countries. Washington (D. C.): World Bank Publications. 1352 p. editors.
- Jiménez M., Dáz-Delgado R. (2015). Towards a Standard Plant Species Spectral Library Protocol for Vegetation Mapping: A Case Study in the Shrubland of Doñana National Park. *ISPRS International Journal of Geoinformation*, 4: 2472 – 2495.
- Kabaria C.W., Molteni F., Mandike R., Chacky F., Noor A.M., Snow R.W., Linard C. (2016). Mapping intra-urban malaria risk using high resolution satellite imagery: a case study of Dar es Salaam. *International Journal of Health Geographics*, 15(26): DOI: 10.1186/s12942-016-0051-y.
- Kaptué A., Hanan N.P., Prihodko L. (2013). Characterization of the spatial and temporal variability of surface water in the Soudan-Sahel region of Africa. *Journal of Geophysical Research: Biogeosciences*, 118: 1472 – 1483.
- Kapwata T., Gebreslasie M.T. (2016). Random forest variable selection in spatial malaria transmission modelling in Mpumalanga Province, South Africa. *Geospatial Health*, 11(3): DOI: 10.4081/gh.2016.434.



- Katole S.B., Kumar P., Patil R.D. (2013). Environmental pollutants and livestock health: A review. *Veterinary Research International*, 1(1): 1 – 13.
- Kauffman C., Briegel H. (2004). Flight performance of the malaria vectors *Anopheles gambiae* and *Anopheles atroparvus*. *Journal of Vector ecology*, 29: 140 – 153.
- Keiser J., Castro D., Maltese M.(2005). Effect of irrigation and large dams on the burden of malaria on a global and regional scale. *American Journal of Tropical Medicine and Hygiene*, 72: 392–406.
- Kellndorfer J., Walker W., Pierce L., Dobson C., Fites J.A., Hunsaker C., Vona J., Clutter M. (2004). Vegetation height estimation from Shuttle Radar Topography Mission and National Elevation Datasets. *Remote Sensing of Environment*, 339 – 358.
- Kelly A.H., Lezaun J. (2013). Walking or Waiting? Topologies of the Breeding Ground in Malaria Control. *Science as Culture*, 22(1): 86 – 107.
- Kerle N. (2010). Satellite-based damage mapping following the 2006 Indonesia earthquake—How accurate was it? *International Journal of Applied Earth Observation and Geoinformation*, 12(6): 466 – 476.
- Komen K. (2017). Could Malaria Control Programmes be Timed to Coincide with Onset of Rainfall? *EcoHealth*, 14(2): 259 – 271.
- Kong J., Sun X., Wong D.W., Chen Y., Yang J., Yan Y., Wang L. (2015). A Semi-Analytical Model for Remote Sensing Retrieval of Suspended Sediment Concentration in the Gulf of Bohai, China. *Remote Sensing*, 7(5): 5373 – 5397.
- Krefis A.C., Schwarz N.G., Nkrumah B., Acquah S., Loag W., Oldeland J., Sarpong N., Adu-Sarkodie Y., Ranft U., May J. (2011). Spatial Analysis of Land Cover Determinants of Malaria Incidence in the Ashanti Region, Ghana. *PLOS ONE*, 6(3):e17905. DOI: 10.1371/journal.pone.0017905.
- Kross A., McNairn H., Lapen D., Sunohara M., Champagne C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied earth Observation and Geoinformation*, 34: 235 – 248.
- Kulkarni A.V., Singh S.K., Mathur P., Mishra V.D. (2006). Algorithm to monitor snow cover using AWiFS data of RESOURCESAT-1 for the Himalayan region. *International Journal of Remote Sensing*, 27(12): 2449 – 2457.
- Lambert M., Waldner F., Defourny P. (2016). Cropland Mapping over Sahelian and Sudanian Agrosystems: A Knowledge-Based Approach Using PROBA-V Time Series at 100-m. *Remote Sensing*, 8(3), 232. DOI: 10.3390/rs8030232.



- Liu T., Zhao X., Shen H., Hu H., Huang W., Fang J. (2016). Spectral feature differences between shrub and grass communities and shrub coverage retrieval in shrub-encroached grassland in Xianghuang Banner, Nei Mongol, China. *Chinese Journal of Plant Ecology*, 40(10): 969 – 979.
- Lopez A.D., Mathers C.D., Ezzati M., Jamison D.T., Murray C.J. (2006). Global and regional burden of disease and risk factors, 2001: Systematic analysis of population health data. *Lancet*, 367: 1747–1757.
- Lopez-Granados F. (2011). Weed detection for site-specific weed management: mapping and real-time approaches. *Weed Research*, 51(1): 1 – 11.
- Lu D., Weng Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5): 823 – 870.
- Lu D., Weng Q. (2004). Spectral mixture analysis of the urban landscapes in Indianapolis with Landsat ETM+ imagery. *Photogrammetric Engineering and Remote Sensing*, 70: 1053–1062.
- Lunde T.M., Boyoh M.N., Lindtjørn B. (2013). How malaria models relate temperature to malaria transmission. *Parasites and Vectors*, 6(20): DOI: 10.1186/1756-3305-6-20
- Lysenko A.J., Semashko I.N. (1968). Geography of malaria. A medico-geographic profile of an ancient disease. *Itogi Nauk Medical Geography*, 25(146).
- Mabogo D.E.N. (1990). The Ethnobotany of the Vhavenda. M.Sc. Thesis. University of Pretoria.
- Machault V., Gadiaga L., Vignolles C., Jarjaval F., Bouzid S., Sokhna C., Lacaux J., Trape J., Rogier C., Pagès F. (2009). Highly focused anopheline breeding sites and malaria transmission in Dakar. *Malaria Journal*, 8(138): DOI: 10.1186/1475-2875-8-138.
- Machault V., Vignolles C., Pages F., Gadiaga L., Toure Y.M., Gaye A., Solhna C., Trape J., Lacaux J., Rogier C. (2012). Risk Mapping of *Anopheles gambiae* s.l. Densities Using Remotely-Sensed Environmental and Meteorological Data in an Urban Area: Dakar, Senegal. *PLOS ONE*, 7(11): e50674. DOI: 10.1371/journal.pone.005067.
- Madonsela S., Cho M.A., Ramoelo A., Mutanga O. (2017). Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, 133: 116 – 127.

- Malahlela O.E., Olwoch J.M., Adjorlolo C. (2018). Evaluating Efficacy of Landsat-Derived Environmental Covariates for Predicting Malaria Distribution in Rural Villages of Vhembe District, South Africa. *EcoHealth*, 15(1): 23 – 40.
- Minale A.S., Alemu K. (2018). Mapping malaria risk using geographic information systems and remote sensing: The case of Bahir Dar City, Ethiopia. *Geospatial Health*, 13(1): DOI: 10.4081/gh.2018.660.
- Mishra D., Narumalani S., Lawson M., Rundquist D. (2004). Bathymetric Mapping Using IKONOS Multispectral Data. *GIScience and Remote Sensing*, 41(4): 301 – 321.
- Mohamed M.F. (2015). Satellite data and real time stations to improve water quality of Lake Manzalah. *Water Science*, 29(1): 68 – 76.
- Mohan V.R., Naumova E.N. (2014). Temporal changes in land cover types and the incidence of malaria in Mangalore, India. *International Journal of Biomedical Research*, 5(8): 494 – 498.
- Montosi E., Manzoni S., Porporato A., Montanari A. (2012). An eco-hydrologic model of malaria outbreaks. *Hydrological Earth Systems Science*, 9: 2831 – 2854.
- Müller G.C., Beier J.C., Traore S.F., Toure M.B., Traore M.M., Bah S., Doumbia S., Schlein Y. (2010). Field experiments of *Anopheles gambiae* attraction to local fruits/seedpods and flowering plants in Mali to optimize strategies for malaria vector control in Africa using attractive toxic sugar bait methods. *Malaria Journal*, 9(262): DOI: 10.1186/1475-2875-9-262.
- Mushinzimana E., Munga S., Minakawa N., Li L., Feng C., Bian L., Kitron U., Schmidt C., Beck L., Zhou G., Githeko A.K., Yan G. (2006). Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. *Malaria Journal*, 5:13. DOI: 10.1186/1475-2875-5-13.
- Mutero C.M., Ng'ang'a P.N., Wekoyela P., Githure J., Konradsen F. (2004). Ammonium sulphate fertiliser increases larval populations of *Anopheles arabiensis* and *culicine* mosquitoes in rice fields. *Acta Tropica*, 89(2): 187 – 192.
- Mwangangi J.M., Mbongo C.M., Muturi E.J., Nzovu J.G., Yan G., Minakawa N., Novak R., Beier J.C. (2007). Spatial distribution and habitat characterization of *Anopheles* larvae along the Kenyan coast. *Journal of Vector Borne Diseases*, 44(1): 44 – 51.
- Nawar S., Buddenbaum H., Hill J., Kozak J., Mouazen A.M. (2016). Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil and Tillage Research*, 155: 510 – 522.

- Ndenga B.A., Simbauni J.A., Mbugi J.P., Githeko A.K., Fillinger U. (2011). Productivity of Malaria Vectors from Different Habitat Types in the Western Kenya Highlands. *PLOS ONE*, 6(4): e19473. DOI: 10.1371/journal.pone.0019473.
- Ngarakana-Gwarisa E.T., Bhunu C.P., Mashonjowa E. (2014). Assessing the impact of temperature on malaria transmission dynamics. *Afrika Matematika* , 25(4): 1095 – 1112.
- Nikolaos S. (1988). Visual and digital classification of Landsat TM data for soil, physiography and land use mapping in Axios alluvial plain, Thessaloniki, Greece. *Geocarto International*, 3(4): 55 – 65.
- Noor N., Lockaby B.G., Kalin L. (2015). Larval development of *Culex quinquefasciatus* in water with low to moderate pollution levels. *Journal of Vector Ecology*, 40(2): 208 – 220.
- Omumbo J.A., Hay S.I., Snow R.W., Tatem A.J., Roers D.J. (2005). Modeling malaria risk in East Africa at high spatial resolution. *Tropical Medicine and International Health*, 10(6): 557 – 566.
- Orych A., Walczykowski P., Dabrowski R., Kutyna E. (2013). Using plant spectral response curve in detecting plant stress. *Ecological Questions*, 17: 67 – 74.
- Oyewole I.O., Momoh O.O., Anyasor G.N., Ogunnowo A.A., Ibadapo C.A., Oduola O.A., Obansa J.B., Awolola T.S. (2009). Physico-chemical characteristics of *Anopheles* breeding sites: Impact on fecundity and progeny development. *African Journal of Environmental Science and Technology*, 3(12): 447 – 452.
- Paul P., Kangalawe R.Y.M., Mboera L.E.G. (2018). Land-use patterns and their implication on malaria transmission in Kilosa District, Tanzania. *Tropical Diseases, Travel Medicine and Vaccines* 4(6): DOI: 10.1186/s40794-018-0066-4.
- Peng D., Jiang Z., Huete A.R., Ponce-Campos G.E., Nguyen U., Luvall J.C. (2013). Response of Spectral Reflectances and Vegetation Indices on Varying Juniper Cone Densities. *Remote Sensing*, 5: 5330 – 5345.
- Pereira O.J.R., Mellfi A.J., Montes C.R. (2017). Image fusion of Sentinel-2 and CBERS-4 satellites for mapping soil cover in the Wetlands of Pantanal. *International Journal of Remote Sensing*, 8(2): 148 – 172.
- Prabhakara K., Hively W.D., McCarthy G.W. (2015). Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, United States. *International Journal of Applied Earth Observation and Geoinformation*, 39: 88 - 102.





- Rahman A., Kogan F., Roytman L., Goldberg M., Guo W. (2011). Modelling and prediction of malaria vector distribution in Bangladesh from remote-sensing data. *International Journal of Remote Sensing*, 32(5): 1233 – 1251.
- Rahman A., Krakauer N., Roytman L., Goldberg M., Kogan F. (2010). Application of Advanced Very High Resolution Radiometer (AVHRR)-based Vegetation Health Indices for Estimation of Malaria Cases. *American Journal of Tropical Medicine and Hygiene*, 82(6): 1004-1009.
- Ramoelo A., Cho M.A. (2018). Explaining Leaf Nitrogen Distribution in a Semi-Arid Environment Predicted on Sentinel-2 Imagery Using a Field Spectroscopy Derived Model. *Remote Sensing*, 10(2): DOI: 10.3390/rs10020269.
- Ramoelo A., Dzikiti S., van Deventer H., Maherry A., Cho M.A., Gush M. (2015). Potential to monitor plant stress using remote sensing tools. *Journal of Arid Environments*, 113: 134 – 144.
- Rebelo L.M., Finlayson C.M., Nagabhatla N. (2009). Remote sensing and GIS for wetland inventory, mapping and change analysis. *Journal of Environmental Management*, 90(7): 2144 – 2153.
- Ricotta E.E., Frese S.A., Choobwe C., Louis T.A., Shiff C.J. (2014). Evaluating local vegetation cover as a risk factor for malaria transmission: a new analytical approach using ImageJ. *Malaria Journal*, 13(94): DOI: 10.1186/1475-2875-13-94.
- Romero E., Vecchia G.D., Jommi C. (2011). An insight into the water retention properties of compacted clayey soils. *Géotechnique*, 61(4): 313 – 328.
- Roosjen P.P.J., Brede B., Suomalainen J.M., Bartholomeus H.M., Kooistra L., Clevers J.G.P.W. (2018). Improved estimation of leaf area index and leaf chlorophyll content of a potato crop using multi-angle spectral data – potential of unmanned aerial vehicle imagery. *International Journal of Applied earth Observation and Geoinformation*, 66: 14 – 26.
- Rubio-Palis Y., Bevilacqua M., Medina D.A., Moreno J.E., Cardenas L., Sanchez V., Estrada Y., Anaya W., Martinez Á. (2013). Malaria entomological risk factors in relation to land cover in the Lower Caura River Basin, Venezuela. *Memórias do Instituto Oswaldo Cruz*, 108(2): 220 – 228.
- Sant'ana A.L., Roque R.A., Eiras A.E. (2014). Characteristics of Grass Infusions as Oviposition Attractants to *Aedes (Stegomyia)* (Diptera: Culicidae). *Journal of Medical Entomology*, 43(2): 212 – 220.

- Sarmah S., Jia G., Zhang A. (2018). Satellite view of seasonal trends and controls in South Asia. *Environmental Research Letters*, 13: 034026.doi: 10.1088/1748-9326/aaa866.
- Sattler M.A, Mtasiwa D., Kiama M., Premji Z., Tanner M., Killeen G.F., Lengeler C. (2005). Habitat characterization and spatial distribution of *Anopheles* sp. mosquito larvae in Dar es Salaam (Tanzania) during an extended dry period. *Malar J.* 2005, 4(1):4.
- Schneider A., Friedl M.A., Potere D. (2010). Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sensing of Environment*, 114(8): 1733 – 1746.
- Scudiero E., Skaggs T.H., Corwin D.L. (2015). Regional-scale soil salinity assessment using Landsat ETM + canopy reflectance. *Remote Sensing of Environment*, 169: 335 – 343.
- Sellers P.J. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, 6: 1335 – 1372.
- Service M.W. (1971). The daytime distribution of mosquitoes resting in vegetation. *Journal of Medical Entomology*, 8(3): 271 – 278.
- Seyoum A., Pålsson K., Kung'a S., Kabiru E.W., Lwande W., Killeen W.F., Hassanali A., Knots B.G.J. (2002). Traditional use of mosquito-repellent plants in western Kenya and their evaluation in semi-field experimental huts against *Anopheles gambiae*: ethnobotanical studies and application by thermal expulsion and direct burning. *Journal of Urology*, 96(3): 225 – 231.
- Sha Z., Zhong J., Bai Y., Tan X., Li J. (2016). Spatio-temporal patterns of satellite-derived grassland vegetation phenology from 1998 to 2012 in Inner Mongolia, China. *Journal of Arid Land*, 8(3): 462 – 477.
- Sharma V.P. (2014). Water, mosquitoes and malaria. In: Singh P., Sharma V. (eds) *Water and Health*. Springer, New Delhi. Springer, New Delhi, pp 155 – 168.
- Sharma V.P. (1996). Re-emergence of malaria in India. *Indian Journal of Medical Research*, 103: 26 – 45.
- Shiff C.J., Stoyanov C., Choobwe C., Kamanga A., Mukonka V.M. (2013). Measuring malaria by passive case detection: a new perspective based on Zambian experience. *Malaria Journal*, 12(120): DOI:10.1186/1475-2875-12-120.





- Sidike A., Zhao S., Wen Y. (2014). Estimating soil salinity in Pingluo County of China using QuickBird data and soil reflectance spectra *International Journal of Applied Earth Observation and Geoinformation*, 26: 156 – 175.
- Snow R.W., Noor A.M. (2015). Malaria risk mapping in Africa: The historical context to the Information for Malaria (INFORM) project. Working Paper in support of the INFORM Project funded by the Department for International Development and the Wellcome Trust , Nairobi, Kenya 3rd June 2015.
- Soleimani-Ahmadi M., Vatandoost H., Hanafi-Bojd A.A., Zare M., Safari R., Mojahedi A., Poorahmad-Garbandi F. (2013). Environmental characteristics of *Anopheles* mosquito larval habitats in a malaria endemic area in Iran. *Asian Pacific Journal of Tropical Medicine*, 6(7): 510 – 515.
- Soleimani-Ahmadi M., Vatandoost H., Zare M. (2014). Characterization of larval habitats for *Anopheles* mosquitoes in a malarious area under elimination program in the southeast of Iran. *Asian Pacific Journal of Tropical Biomedicine*, 4(supp. 1): S73 – S80
- Sothe C., de Almeida C.M., Liesenberg V., Schimanski M.B. (2017). Evaluating Sentinel-2 and Landsat-8 Data to Map Successional Forest Stages in a Subtropical Forest in Southern Brazil. *Remote Sensing*, 9(838): DOI: 10.3390/rs9080838.
- Sousa A., Andrade F., Felix A., Jurado V., Leon-Botubol A., Garcia-Murillo P., Garcia-Barron L., Morales J. (2009). Historical importance of wetlands in malaria transmission in southwest of Spain. *Limnetica*, 28(2): 283 – 300.
- Sumfleth K., Duttman R. (2008). Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecological Indicators*, 8(5): 485 – 501.
- Sun P., Wahbi S., Tsonev T., Haworth M., Liu S., Centrito M. (2014). On the Use of Leaf Spectral Indices to Assess Water Status and Photosynthetic Limitations in *Olea europaea* L. during Water-Stress and Recovery. *PLOS ONE*, 9(8): e105165. 10.1371/journal.pone.0105165.
- Tanser F.C., Sharp B., le Sueur D. (2003). Potential effect of climate change on malaria transmission in Africa. *The Lancet*, 362(9398): 1792 – 1798.
- Tchigossou G., Akoton R., Yessoufou A., Djegbe I., Zeukeng F., Atoyebi S.M., Tossou E., Moutairou K., Djouaka R. (2017). Water source most suitable for rearing a sensitive malaria vector, *Anopheles funestus* in the laboratory. *Wellcome Open Research*, 2(109): DOI:10.12688/wellcomeopenres.12942.2.

- Terashima I., Fujita T., Inoue T., Chow W.S., Oguchi R. (2009). Green Light Drives Leaf Photosynthesis More Efficiently than Red Light in Strong White Light: Revisiting the Enigmatic Question of Why Leaves are Green. *Plant and Cell Physiology*, 50(4): 684 – 697.
- Thapa R.B., Itoh T., Shimada M., Watanabe M., Takeshi M., Shiraishi T. (2014). Evaluation of ALOS PALSAR sensitivity for characterizing natural forest cover in wider tropical areas. *Remote Sensing of Environment*, 155: 32 – 41.
- Thomas C.J., Davies G., Dunn C.E. (2004). Mixed picture for changes in stable malaria distribution with future climate in Africa. *Trends in Parasitology*, 20(5): 216 – 220.
- Tian S., Zhang X., Tian J., Sun Q. (2016). Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. *Remote Sensing*, 8(11): DOI: 10.3390/rs8110954.
- Valle D., Zaitchik B., Feingold B., Spangler K., Pan W. (2013). Abundance of water bodies is critical to guide mosquito larval control interventions and predict risk of mosquito-borne diseases. *Parasites and Vectors*, 6(179): DOI: 10.1186/1756-3305-6-179.
- van der Werff H., van der Meer F. (2015). Sentinel-2 for Mapping Iron Absorption Feature Parameters. *Remote Sensing*, 7(10). 12635 – 12653.
- Verstraeten W.W., Veroustraete F., van der Sande C., Grootaers I., Feyen J. (2006). Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests. *Remote Sensing of Environment*, 101(3): 299 – 314.
- Vijith H. (2007). Groundwater potential in the hard rock terrain of Western Ghats: A case study from Kottayam district, Kerala using resourcesat (IRS-P6) data and GIS techniques. *Journal of the Indian Society of Remote Sensing*, 35(163): DOI: 10.1007/BF02990780.
- Viscarra Rossel R.A., Bherens T., Ben-Dor E., Brown D.J., Demattê J.A.M., Shepherd K.D., Shi Z., Stenberg B., Stevens A., Adamchuk V., Aichi H., Barthès B.G., Bartholomeus H.M., Bayer A.D., Bernoux M., Böttcher K., Brodský L., Du C.W., Ju W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155: 198 – 230.
- Vos R.J.J., Hakvoort H.M., Jordans R.W.J., Ibelings B.W. (2003). Multiplatform optical monitoring of eutrophication in temporally and spatially variable lakes. *Science of the Total Environment*, 312: 221 – 243.



- Wagner W., Naeimi V., Scipal K., de Jeu R., Martinez-Fernandez J. (2007). Soil moisture from operational meteorological satellites. *Hydrogeology Journal*, 15(1): 121 – 131.
- Wang B., Ono A., Muramatsu K., Fujiwara N. (1999). Automated detection and removal of clouds and their shadows from Landsat TM images. *IEICE Transactions on Information and Systems*, E82- D, 453–60.
- Wang L., Qu J.J. (2007). NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophysical Research Letters*, 34(20): L20405. DOI: 0.1029/2007GL03102.
- Wang X., Wang K., Zhou B. (2011). Object-Based Classification of IKONOS Data for Endemic *Torreya* Mapping. *Procedia Environmental Sciences*, 1887 – 1891.
- Wardrop N.A., Barnett A.G., Atkinson J., Clements A.C.A. (2013). *Plasmodium vivax* malaria incidence over time and its association with temperature and rainfall in four counties of Yunnan Province, China. *Malaria Journal*, 12(452): DOI: 10.1186/1475-2875-12-452.
- Waudby H.P., Petit S. (2015). Disintegration of cattle hoof prints in cracking-clay soils of the arid South Australian stony plains region during a wet period. *South Australian Geographical Journal*, 113: 5 – 12.
- Weiss D.J., Mappin B., Dalrymple U., Bhatt S., Cameron E., Hay S.I., Gething P.W. (2015). e-examining environmental correlates of *Plasmodium falciparum* malaria endemicity: a data-intensive variable selection approach
- Wirion C., Bauwens W., Verbeiren B. (2017). High resolution modelling of the urban hydrological response. *Joint Urban Remote Sensing Event (JURSE) Conference Proceedings*, IEEE, Dubai, United Arab Emirates. 6-8 March 2017. DOI: 10.1109/JURSE.2017.7924555.
- Wright J.P., Ashworth A.J., Allen F.L. (2016). Organic substrate, clay type, texture and water influence on NIR carbon measurements. *Geoderma*, 261: 36 – 43.
- Wu C., Jacobson A.R., Laba M., Baveye P.C. (2009). Accounting for surface roughness effects in the near-infrared reflectance sensing of soils. *Geoderma*, 152(1-2): 171 – 180.
- Yan J., Zhou W., Han L., Qian Y. (2018). Mapping vegetation functional types in urban areas with WorldView-2 imagery: Integrating object-based classification with phenology. *Urban Forestry and Urban Greening*, 31: 230 – 240.

- Yang C., Ren X., Huang H. (2012). The vegetation damage assessment of the Wenchuan earthquake of May 2008 using remote sensing and GIS. *Natural Hazards*, 62(1): 45 – 55.
- Ye-Ebiyo Y., Pollack R.J., Kiszewski A., Spielman A. (2003). Enhancement of development of larval *Anopheles arabiensis* by proximity to flowering maize (*Zea mays*) in turbid water and when crowded. *American Journal of Tropical Medicine and Hygiene*, 68(6): 748 – 752.
- Zhang L., Sun X., Wu T., Zhang H. (2015). An Analysis of Shadow Effects on Spectral Vegetation Indexes Using a Ground-Based Imaging Spectrometer. *IEEE Geoscience and Remote Sensing Letters*, 12(11): 2188 – 2192.
- Zhang M., Dong Q., Tang J., Song Q. (2010). Evaluation of the retrieval of total suspended matter concentration in Taihu Lake, China from CBERS-02B CCD. *Chinese Journal of Oceanology and Limnology*, 28(6): 1316 – 1322.
- Zhang X., Friedl M.A., Schaaf C.B., Strhler A.H., Liu Z. (2005). Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments. *Journal of Geophysical Research: Atmospheres*, 110(D12): DOI: 10.1029/2004JD005263.
- Zhang Y., Wang C., Wu J., Qi J., Salsa W.A. (2009). Mapping paddy rice with multitemporal ALOS/PALSAR imagery in southeast China. *International Journal of Remote Sensing*, 30(23): 6301 – 6315.
- Zheng G., DiGiacomo P.M. (2017). Remote sensing of chlorophyll-a in coastal waters based on the light absorption coefficient of phytoplankton. *Remote Sensing of Environment*, 201L 331 – 341.



## CHAPTER 3<sup>2</sup>:

# Integrating geostatistics and remote sensing for mapping the spatial distribution of cattle hoofprints in relation to malaria vector control

Oupa E. Malahlela <sup>a,b\*</sup>, Adjorlolo C <sup>c.</sup>, Olwoch J.M. <sup>a,d</sup>, Kganyago L.M <sup>b</sup>, Mashalane M.J. <sup>b</sup>

<sup>a</sup> *Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa*

<sup>b</sup> *South African National Space Agency (SANSA), Earth Observation Directorate, Pretoria 0001, South Africa*

<sup>c</sup> *New Partnership for Africa's Development (NEPAD) Agency, 230 15<sup>th</sup> Road, Midrand, South Africa*

<sup>d</sup> *Southern African Science Service Center for Climate Change and Adaptive Land Management (SASSCAL), Windhoek 91100, Namibia*

***This paper relates to objectives 1 of the thesis.***

---

<sup>2</sup>This chapter is based on the manuscript titled "*Integrating geostatistics and remote sensing for mapping the spatial distribution of cattle hoofprints in relation to malaria vector control*" **International Journal of Remote Sensing**. <https://doi.org/10.1080/01431161.2019.1584688>

## Abstract

Globally, malaria is still a persistent health problem affecting more than 200 million people. With about 90% of malaria cases occurring in Sub-Saharan Africa, it becomes imperative to understand the environmental factors contributing to malaria vector proliferation. The cattle hoofprints are known to be some of the productive breeding sites for *Anopheles (An.) arabiensis* and *An. fenestus* in Southern and East African countries. Therefore, this chapter aimed at testing the potential of integrating field data and Sentinel-2 satellite imagery for mapping cattle hoofprint distribution in the Vhembe District, South Africa. The purpose was to improve the predictability of mosquito breeding sites in the study area by using field point dataset and Sentinel-2 data. Due to the difficulty of sampling all locations at the study area, the spatial interpolation was employed to create continuous surfaces of cattle hoofprints, using limited sampled point observations. The sampled point observations were then correlated with Sentinel-derived variables for predicting cattle hoofprints at unsampled locations. The ordinary kriging (OK), co-kriging (CK) and step-wise multiple linear regression (SMLR) were used due to their ability to incorporate both field point data and ancillary datasets. The CK was the best performing interpolation method, with  $R^2 = 0.69$  for validation dataset ( $n = 33$ ), compared to OK ( $R^2 = 0.57$ ) and SMLR ( $R^2 = 0.25$ ). The resulting co-kriging semivariogram shows that the combination of field data and remote sensing dataset improves the prediction accuracy of cattle hoofprint distribution. Findings from this chapter demonstrated that the interpolation error for estimating cattle hoofprints/100 m<sup>2</sup> can be minimized greatly by using CK (RMSE = 0.2; MAD = 0.04) than with both OK (RMSE = 2.39; MAD = 2.11) and SMLR (RMSE = 5.20; MAD = 4.55) methods. Furthermore, the results from this chapter indicate that there is high number of cattle hoofprints in malaria-prone areas at the study site than in the malaria free areas. Studies such as this provide the platform for developing operational platform for long-term monitoring of areas susceptible to malaria, risks and control management.

**Keywords:** Vhembe District Municipality; geostatistics; cattle hoofprints; malaria; S-2

---

### 3.1. Introduction

Malaria continues to be one of the major public health concerns globally. Although the infection is preventable and treatable, a large population remains without access to malaria prevention and treatment, with most cases and deaths going unregistered and unreported (WHO, 2015a). By the end of 2015, the malaria incidence rate fell by 37% while mortality rate fell by approximately 60% globally (Sewe *et al.*, 2017; WHO, 2015a). The reduction in new infections and mortality is largely attributed to the malaria prevention strategies put in place at local, regional and global scale (WHO, 2015b). Despite various intervention strategies globally,

malaria continues to exert pressure on local and international health systems. In order to avert impacts of malaria on millions of people worldwide it is imperative to increase the understanding of spatial distribution of vector habitats for targeted disease control. Consistent and accurate quantification of the vector breeding habitats could be imperative for rolling back malaria in Africa.

Previous studies have shown that the primary malaria vectors (*An.arabiensis* and *An.gambiae* for example) readily rest and feed outside and inside houses (Mayagaya *et al.*, 2015; Braack *et al.*, 2015). However, elsewhere in East Africa, Highton *et al.* (1979) and Joshi *et al.* (1975) have showed that *An. arabiensis* species has a tendency of frequently occurring outdoors than indoors more than 2.2 of the times. The presence of livestock (such as cattle) close to households affects the risk of *Anophele* malaria transmission to humans (Franco *et al.*, 2006), especially in rural areas. Livestock may impact on the malaria transmission mainly in two ways, (i) acting as the zoophylactic agents (alternative host species for biting mosquitoes), and (ii) zoopotential (attracting more malaria vectors). There have been studies conducted to suggest the potential use of cattle livestock as alternative host for *Anophele* species, in order to divert malaria transmission from humans (Franco *et al.*, 2014; WHO., 1982; Habtewold, 2004). On the latter case, cattle livestock has shown to attract more than 90% of the collected *An. arabiensis* in Moshi, Northern Tanzania (Mahande *et al.*, 2007). The abundance of cattle and other domestic animals may result in increased vector densities thus increasing malaria transmission (Bouma and Rowland, 1995). This is mainly attributed to the heat and odour cues emitted by cattle as well as the physical disturbances created by animals such as hoofprints and puddles near households (Mayagaya *et al.*, 2015; Iwashita *et al.*, 2014). The presence of cattle in marshy areas results in the creation of hoofprints that potentially offer ideal conditions for mosquito breeding (Lobo, 2010). These ideal micro breeding habitats should be taken into consideration when planning the outdoor control strategies, especially in respect to *An. arabiensis* and *An. fenestus*.

In Southern Africa and particularly South Africa, Zimbabwe and Mozambique, malaria is transmitted through the bite of female *Anopheles arabiensis* (*An. arabiensis*) and *An. fenestus* mosquitoes (Brooke *et al.*, 2015; Sande *et al.*, 2015). It has been documented that these vectors breed in small ambient aquatic habitats within village areas, including the cattle hoofprints (Mayagaya *et al.*, 2015; Davies, 2016; Charlwood *et al.*, 2013). These habitats are usually temporary and tend to be concentrated in predominantly disturbed landscapes in rural environments (Sinka *et al.*, 2010). Although the *An. arabiensis/fenestus* habitat preferences are known, the information on spatial distribution of cattle hoofprints is very limited or scanty



both globally and regionally. Providing spatial information on cattle hoof print will essentially enhance the outdoor vector control strategies especially for the more exophylic species such as *An. arabiensis* (Mahande *et al.*, 2007). In addition, detailed assessment of the correlation between cattle hoofprint distribution and malaria prevalence is poorly understood. Therefore, mapping the distribution of cattle hoofprints across landscape is crucial for informing the efforts aimed at preventing malaria and for treating outdoor potential *An. arabiensis/fenestus* habitats found at close proximity to rural households. This is particularly important for *An. arabiensis* which is zoophilic and prefers to rest outdoors. Moreover, this will also assist in enhancing the understanding of the environments surrounding the *An. arabiensis/fenestus* potential breeding habitats, and thus contributing to reductions in malaria incidence rates.

A number of environmental conditions have been established as key contributors to the occurrence and distribution of malaria breeding habitats in various parts of the world. For example, physiochemical factors such as water, salinity and vegetation greenness have been shown to correlate with *Anophele* larvae in pools and cattle hoofprints (Pfaehler *et al.*, 2006; Gimning *et al.*, 2001; Hurd, 2014). Previous entomological studies have found that the female *Anophele* mosquito requires temporary surface water for laying eggs, which usually take 2-3 days before hatching (Zhou *et al.*, 2010). On the other hand, soil plays a significant role in the occurrence and distribution of cattle footprint across the landscape. A study by Huang *et al.* (2006) showed that *An. gambiae* preferred to deposit more eggs on bare, wet soils than grass-covered soil. Potential habitats is bare wet soil may be linked to a number of cattle hoofprints created as a result of physical disturbance (McLaughlin and Vidrine, 1987). Vegetation also correlated with the oviposition of *An.* species in micro-habitats (cattle hoofprints and puddles) and as such exerts a major impact on the site selection for *Anophele* mosquitoes (Huang *et al.*, 2006). The relationship between cattle density and vegetation type is also important in mapping cattle hoofprints by dictating the hoofprint accounting, and such a relationship has been documented elsewhere (Moleele and Perkins, 1998). Characterizing the soil, vegetation, and water/moisture within the confinement of cattle hoofprint distribution is critical for targeted mosquito larval control strategies.

In order to assess the ecological conditions associated with soil (as in the case of cattle hoofprint distribution) scientists have successfully adopted the use of spatial interpolation techniques such as Kriging (Shit *et al.*, 2016; Deis *et al.*, 2017; Reza *et al.*, 2017). In this way, spatial distribution of certain soil characteristics incorporating detailed information are effectively mapped and presented more accurately. Moreover, interpolation methods such as kriging are mostly preferred due to their ability to provide estimates of unobserved locations



of variables, based on their weighted averages of adjacent observed sites within given study site (Setianto and Triandini, 2013). Kriging has also been applied for mapping vegetation characteristics and soil-water parameters (Mutanga and Rugege, 2006; Adjorlolo and Mutanga, 2013; Kambhammettu *et al.*, 2011). However, as a geostatistical tool kriging mainly offers the capability to derive predicted spatial surfaces using observed point locations, and does not incorporate ancillary data sources. Considering the dynamics of cattle hoofprints as one of the crucial *Anopheles* breeding habitats, there is a need to integrate to utilize methods and data that could improve mapping of such hoofprints in an area. Co-kriging is an extension of ordinary kriging and has shown to perform better than ordinary kriging method due to its ability to accommodate cross-correlation and spatial dependency of the field and ancillary datasets (Adjorlolo and Mutanga, 2013). The use of co-kriging and ancillary data has shown to improve the prediction accuracy of parameters in soil-water-vegetation nexus by reducing prediction errors associated with spatial interpolation (Carter *et al.*, 2011; Adjorlolo and Mutanga, 2013; Mutanga and Rugege, 2006; Song *et al.*, 2014). In order for a co-kriging interpolation to successfully predict cattle hoofprints, there is a need to incorporate suitable ancillary data when interpolating.

Remote sensing data are recognized as some of the readily available datasets for effective landscape mapping by way of spatial interpolation (Setianto and Triandini, 2013; López-Granados *et al.*, 2005; Arieira *et al.*, 2010). Remote sensing data cover extensive areas and capture spatial and spectral characteristics of landscape features including cattle hoofprints. High densities of cattle hoofprints tend to alter the soil physiochemical characteristics (Pietola *et al.*, 2005), which impacts on the spectral response of such features by remote sensing satellites. The availability of new earth observation satellites at high spatial and spectral resolutions offers opportunity to improve on the detection of surface features, and could potentially serve as input parameters for co-kriging, thus improving spatial interpolation of cattle hoofprints. One such satellite is the Sentinel-2 (S-2) with 13 multispectral data at 10 m spatial resolution. The S-2 data is freely available and have high repeat cycle. This dataset thus provides capability to estimate malaria vector microhabitats at finer scale while capturing habitat conditions at short intervals. These attributes are key to understanding malaria vector habitat and would aid in the identification of areas with potentially high likelihood of *Anopheline* larvae occurrence (Sattler *et al.*, 2005). The primary goal of this chapter is to integrate field data and remote sensing technology to map the distribution of cattle hoofprints, using the Vhembe District of South Africa as a case study, by means of interpolation technique. The study has the following objectives:

- (i) To compare the performance of kriging, co-kriging and linear regression for interpolating cattle hoofprint using field observation data only and/or integration of field and remote sensing data, and
- (ii) To produce a spatial distribution map of cattle hoofprints using the best-performed interpolation technique.

Here, for the first time, we show the feasibility of integrated ground-based geostatistical methods and remote sensing methods for mapping the distribution of *Anopheles* microhabitats in the form of cattle hoofprints. This study will eventually assist in efforts aimed at eliminating malaria in subtropical regions of Southern Africa, particularly by targeting habitats with high potential of harbouring high densities of *Anopheline* larvae.

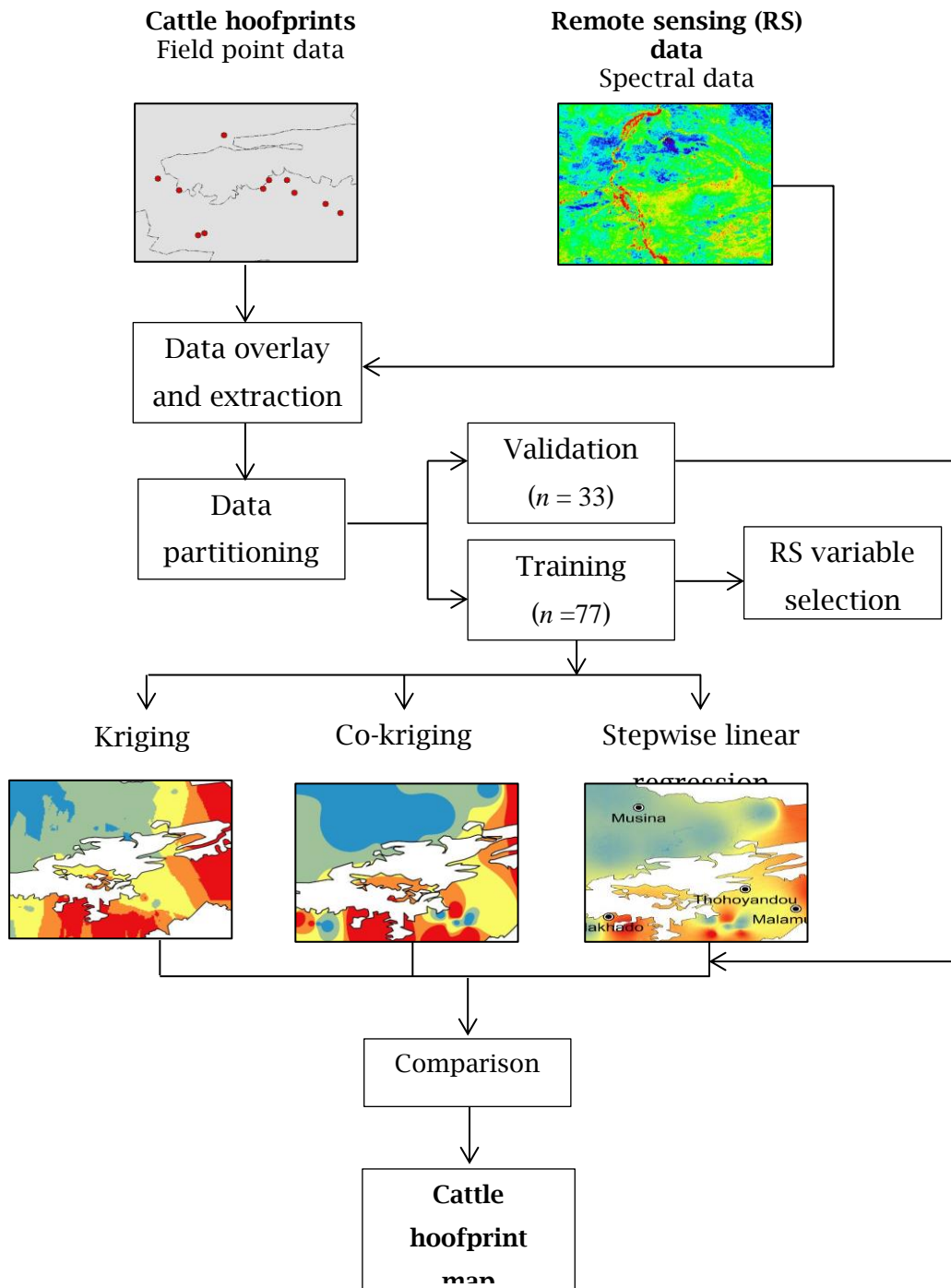
## 3.2. Methods

### 3.2.1. Study area and malaria data

The experimental study was conducted in Vhembe District Municipality (VDM) in Limpopo province of South Africa, as defined in [Malahlela et al. \(2018\)](#). In the VDM – a district comprised largely of Vhavenda people – most household keep cattle livestock, which is perceived by locals as an indication of wealth ([Ramudzuli and Horn, 2014](#)). Figure 3.1 outlines the workflow adopted for mapping the distribution of cattle hoofprints in the Vhembe District Municipality, South Africa.

### 3.2.2. Field data acquisition and transformation

The extensive field data collection was undertaken in April 2017 and again in August 2017. In Vhembe District the month of April corresponds to the peak of malaria transmission season ([Moonasar et al., 2011](#)) while August corresponds to the low transmission period ([Gerritsen et al., 2008](#)). With respect to vegetation, the month of April to November are characterized as dry season months in the southern African savannah biome ([Archibald and Scholes, 2007](#)). The data was collected at various land cover types, including (i) shrublands, (ii) croplands, (iii) grassland, (iv) woodland, and (v) bareland. A 30m×30m plot was randomly selected in largely homogenous areas belonging to any of the land cover type. In each plot, three sub-plots of 10m × 10m (100 m<sup>2</sup>) were sampled to record the soil type data, the number of cattle hoofprints, the vegetation characteristics, soil characteristics and the geographic location. The accounting of cattle hoofprints was done per individual plot print, and where two hoofprints greatly overlapped, a single print was recorded instead.



**Figure 3.1:** Summary of the methodology adopted for the study

The geographic location was recorded using hand-held Garmin eTrex 20™, with the maximum spatial accuracy of 3m. In all locations the total of 110 sampling points ( $N=110$ ) was collected throughout fieldwork.

These field data points were transformed to Gaussian distributions to minimize skewness, since they comprised of ranges between 0 and 18 (Table 3.1). The hoofprints data was

transformed using the logarithmic transformation, which is considered where the coefficient of skewness is greater than 1 (Webseter and Oliver, 2001) and such transformation is given in equation. 3.1:

$$\text{Log}_y = \log(x_i + 1) \quad (3.1)$$

where  $\text{Log}_y$  is log-transformed value of the hoof print of  $x$  at an  $i$ th location. The transformation was done in order for data to meet the linearity assumption (Tabachnick and Fidwell, 1996).

**Table 3.1:**

Descriptive statistics of the measured cattle hoofprints in the study area

	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Std. dev</b>	<b>Mean</b>
Cattle hoofprints	110	0	18	4.44	4.46

### **3.2.3. Remote sensing data acquisition and pre-processing**

S-2 multispectral imagery (MSI) was used for the study. The imagery was acquired for periods between 15 August and 19 October 2017. The S-2 data comprises of 13 spectral bands, in the visible (0.44  $\mu\text{m}$ ) to shortwave infrared region (2.19  $\mu\text{m}$ ). For the purpose of this chapter, only 10 spectral bands were used for the analysis (Table 3.2). Sen2cor atmospheric correction procedure was used to minimize the contribution of atmosphere from the MSI image. Sen2cor uses a large database of look-up tables (LUT) derived using an atmospheric radiative transfer model based on libRadtran1 (Müller-Wilm, 2016). Because the Sentinel images are acquired at ground sampling distances (GSD) of 10m (visible, near infrared), 20m (red edge, shortwave infrared) and 60m (aerosol, water vapour, cirrus) a spatial resampling method was applied on 10 spectral bands used for the study. All the spectral bands used in the study were resampled to 10m using the nearest neighbour resampling method in ENVI 4.7 (Exelis, 2017).

**Table 3.2:**

Multispectral bands of S-2 explored in this chapter

Band	Band center (µm)	GSD (resampled)
Blue	0.49	10m (-)
Green	0.56	10m (-)
Red	0.66	10m (-)
Red edge1	0.70	20m (10m)
Red edge2	0.74	20m (10m)
Red edge3	0.78	20m (10m)
Near infrared	0.84	10m (-)
Red edge4	0.86	20m (10m)
Shortwave infrared1	1.61	20m (10m)
Shortwave infrared1	2.19	20m (10m)

**Table 3.3:**

Spectral indices derived from S-2 that were tested in the current study

Index	Formula	Related to	Reference
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$	Green biomass	<a href="#">Rouse et al. (1974)</a>
Normalized Difference Vegetation Red Edge Index1 (NDVIR <sub>1</sub> )	$NDVIR_1 = \frac{\rho_{NIR} - \rho_{Red\ Edge1}}{\rho_{NIR} + \rho_{Red\ Edge1}}$	Green biomass	<a href="#">Gitelson and Merzlyak (1994)</a>
Normalized Difference Water Index (NDWI)	$NDWI = \frac{\rho_{NIR} - \rho_{SWIR1}}{\rho_{NIR} + \rho_{SWIR1}}$	Moisture, Surface water	<a href="#">Gao (1996)</a>
Simple Ratio Index (SR)	$SRI = \frac{\rho_{NIR}}{\rho_{Red}}$	Moisture	<a href="#">Jackson and Huete (1991)</a>
Ferrous Iron Index (FII)	$FII = \frac{\rho_{Red}}{\rho_{Blue}}$	Red soil exposure, iron concentrations	<a href="#">Clarke (1999)</a>

\* These indices are specially derived from new S-2 red edge bands.

From these bands, the spectral indices sensitive to soil, water, and vegetation were tested in order to understand the correlation between areas of cattle hoofprints and satellite data. The selected spectral indices used in the study are summarized in [table 3.3](#). Because the habitats of *Anopheles* species are largely associated with soil, water and vegetation parameters, a total of five (5) remote sensing spectral indices were selected in the confinement of these parameters. These indices are related to soil bareness, moisture, and vegetation greenness. They were computed based on the possible relationship between cattle density and the vegetation type ([Moleele and Perkins, 1998](#)). The use of spectral indices is advantageous in that they can serve as means to rapidly extract relevant information rapidly and effectively, and their underlying mechanisms are well-understood ([Delegido et al., 2013](#)). They also reduce the effect of multi-collinearity since vegetation, soil and water are and their respective reflectances are spatially correlated both to themselves (auto-correlated) and to another (cross-correlated) ([Myers et al., 1970](#); [Acharya 1999](#)).

One of the most commonly used spectral indices is the normalized difference vegetation index (NDVI) which was calculated as the sum of the difference between near-infrared and the red bands of S-2 as shown in [table 3.3](#). The NDVI is a structural vegetation parameter that is sensitive to green vegetation density and phenology ([Hmimina et al., 2011](#)). Previous studies have shown that NDVI suffers from saturation problems, especially when used to estimate vegetation biophysical and biochemical characteristics ([Malahlela et al., 2014](#); [Mutanga and Skidmore, 2004](#)). In order to compensate for the limitation resulting from the use of NDVI, [Mutanga and Skidmore \(2004\)](#) have recommended that the red edge band (705 nm) since it addresses signal saturation problem common in NDVI. For this reason, the red edge NDVI was included in this study in order to understand indicators of cattle hoofprint as a function of associated vegetation characteristics. On the other hand, [Gao \(1996\)](#) has used the spectral indicator sensitive to changes in water content of the leaves, and has since been used for hydrological mapping such as water body mapping ([Gao et al., 2016](#); [Malahlela, 2016](#)). Additional to these spectral indices, the soil-related ferrous index (FI) ([Clarke, 1999](#)) and simple ratio index ([Jackson and Huete, 1991](#)) were also calculated to relate soil and vegetation with distribution of cattle hoofprints respectively.

#### **3.2.4. Statistical analysis**

Two geostatistical interpolation methods were tested in an attempt to estimate cattle hoofprints in the Vhembe District Municipality. These methods were (i) ordinary kriging (OK), and (ii) co-kriging (CK). In these interpolation methods, the value of variable *Z* at the unsampled location

$k_0$  is estimated based on the data from the surrounding location,  $Z(k_i)$  (Yao *et al.*, 2013), as in equation 3.2.

$$Zk_0 = \sum_{i=1}^n w_i Z(k_i) \quad (3.2)$$

where  $w_i$  is the weight assigned to each  $Z(k_i)$  value and  $n$  is the number of the closest neighbouring sampled cattle hoofprints in each plot.

#### 3.2.4.1. Ordinary Kriging

The OK calculates the values of  $w_i$  by estimating the spatial structure of the variable's distribution represented by sample variogram as,

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [Z(x_i) - Z(x_i + d)]^2 \quad (3.3)$$

where  $x_i$  and  $x_i + d$  are sampling locations separated by the distance  $d$ ,  $n$  represents the number of pairs of observations separated by  $h$ ,  $\gamma(h)$  is the estimated 'experimental' semi-variance value for all pairs at a lag distance  $h$ ,  $Z(x_i)$  and  $Z(x_i + h)$  are the observed values of cattle hoofprint as the corresponding locations. These geostatistical models were performed on randomly selected calibration data points collected in the field.

#### 3.2.4.2. Co-kriging

Co-kriging (CK) is an extension of ordinary kriging that takes into account the spatial cross-correlation between two variables, i.e. the primary variable (cattle hoofprints in our case), and another ancillary variable (in chapter, remote sensing indices). In the current study, the cattle hoofprints (primary variable) have been sampled at few places while remote sensing variables cover the entire study area at every location. The cross-spatial auto-covariance between primary and secondary variables is quantified when these variables are found to have cross-variogram (Isaaks and Srivastava, 1989; Webster and Oliver, 2001). The cross-variogram is computed through the equation 3.4:

$$\gamma_{uv}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z_u(x_i) - z_u(x + h)][z_v(x_i) - z_v(x + h)] \quad (3.4)$$

where  $\gamma_{uv}(h)$  is the cross-semivariance (cross-variogram) between variables  $u$  and  $v$ ,  $n(h)$  is the number of pairs of data locations separated by lag distance  $h$ ,  $z_u$  is the value of variable



$u$  at the location  $x_i$  and  $(x_i + h)$  and  $z_v$  is the data value of variable  $v$  at the same locations (Mutanga and Rugege, 2006).

Similar to OK, the CK predictions of cattle hoofprint distribution is obtained from the equation 3.5:

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (3.5)$$

where  $Z(x_0)$  is the optimal and unbiased estimate of cattle hoofprint,  $\lambda_i$  is the optimum weight selected to minimize estimation variance (Burrough and McDonnell, 1998), and  $Z(x_i)$  are the actual values of cattle hoofprints. The remote sensing variables correlated with cattle hoofprints were thus used as ancillary datasets. The remote sensing variables in final model were correlated to cattle hoofprints using a correlation coefficient ( $R^2$ ). This was done to explicitly understand the contribution of remote sensing data for estimating the cattle hoofprints at higher spatial resolution (10 m).

#### 3.2.4.3. Step-wise multiple linear regression

The step-wise multiple linear regression (SMLR) has been widely used by many researchers to estimate biophysical and biochemical characteristics of soil, water and vegetation by making use of information provided by spectral features (Kokaly and Clark, 1999). The SMLR is given by equation 3.6 as,

$$y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon \quad (3.6)$$

where  $y$  is the predicted cattle hoofprint at  $i$ th location,  $\beta_0$  is the intercept,  $\beta_n$  is the coefficient of  $x_n$  and  $\varepsilon$  is an error term. The calibration model was derived using MASS package embedded in R software (R Development Core Team, 2017), where the model with lowest Akaike's Information Criterion (AIC) was selected through stepwise method as, variables are ranked by their  $P$  value (Fotheringham *et al.*, 2002). Whereas the SMLR can be used to find subsets that best predict responses on dependent variable (Adjorlolo and Mutanga, 2013), model over-fitting could significantly impact on the outcome of prediction. In order to avoid this statistical challenge it is recommended that the number of predictor variables to enter the model should be less than 1/3 the number of observations, and that the number of steps the SMLR should be 10-20 times less than training observations (Mutanga and Rugege, 2006; Cohen *et al.*, 2003). In this chapter, the input variables were less than 1/3 ( $n=15$ ) and seven steps ( $P=7$ ) were achieved before model selection stabilized.



### 3.2.5. Calibration model

From the standard cattle hoofprint dataset ( $N = 110$ ), 70% of the data ( $n = 77$ ) was randomly drawn and used for interpolation and model calibration. This was done to train the predictive model and to primarily assess the correlation between the interpolated ( $y_i$ ) and measured ( $y_o$ ) cattle hoofprints from calibration dataset. Model validation was achieved using independent validation dataset.

### 3.2.6. Validation model

Approximately one-thirds of the data ( $n = 33$ ) was used for validating the interpolation methods and the final predictive model. The root mean square error (RMSE) and the mean absolute deviation (MAD) were calculated using Eqs. 5 and 6 respectively.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - a_i)^2} \quad (3.7)$$

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n (p_i - a_i) \quad (3.8)$$

where  $n$  is the number of values in the validation dataset,  $p_i$  is the predicted value,  $a_i$  is the actual (measured) value.

## 3.3. Results

### 3.3.1. Relationship between cattle hoofprints and S-2 data

Table 4 shows the relationship between cattle hoofprints and S-2 data using Pearson correlation co-efficient. Generally, the spectral data exhibited weak correlations with cattle hoofprints field dataset. Within these weak correlation results, the NDWI showed the highest correlation coefficient of 0.23. The lowest correlation was found between cattle hoofprints distribution and the S-2 green band, at 0.01. All the remote sensing variables (Table 3.4) were used as input candidates in the subsequent stepwise multiple linear regression model. The final regression model parameters used for mapping hoofprint through SMLR are shown in table 3.5.

**Table 3.4:**

The relationship between cattle hoofprints and spectral datasets ( $n = 77$ ).

Variable	$R^2$
Blue	0.05
Green	0.01
Red	0.04
Red edge1	0.14
Red edge2	0.03
Red edge3	0.04 *
NIR	0.10
Red edge4	0.03
SWIR1	0.15
SWIR2	0.17
NDVI	0.13
NDVIr1	0.14
NDWI	0.23
SR	0.15 **
FII	0.02

Note: Correlation coefficient ( $R^2$ ): \* significant level ( $p < 0.05$ )

### 3.3.2. Spatial interpolation of cattle hoofprints

The OK and CK were used to spatially map the distribution of cattle hoofprints from the calibration dataset ( $n = 77$ ). The OK yielded the correlation coefficient of 0.57 ( $n = 33$ ; RMSE = 2.39; MAD = 2.11 prints per 100 m<sup>2</sup>). Table 3.6 shows the results of geostatistical analysis. On the other hand, the CK had the highest correlation co-efficient of and the lowest estimation errors ( $R^2 = 0.69$ ; RMSE = 0.20; MAD = 0.04 prints per 100 m<sup>2</sup>). The SMLR model yielded the lowest correlation with the cattle hoofprint distribution ( $R^2 = 0.25$ ; RMSE = 5.20; MAD = 4.55 prints per 100m<sup>2</sup>). The overall result was that the  $R^2$  of all models was greater than 0.5 indicating that the interpolated data exhibited a good fit with the measured data. The resultant maps obtained from the application of three geostatistical methods tested in the current study are shown in figure 3.2. The CK model used for prediction was significant at  $p < 0.02$ . The

distribution of cattle hoofprints is highly concentrated in the eastern part of the study area compared to the sparse distribution in the west

**Table 3.5:**

The S-2 variables used in the final SMLR model

Variable	Estimate	Standard Error	p-value
<i>Intercept</i>	1.3256	0.3498	0.0003***
Blue	- 0.0011	0.0004	0.0076**
Red edge1	0.0009	0.0003	0.0045**
Red edge2	- 0.0029	0.0010	0.0042**
Red edge3	0.0025	0.0009	0.0099**
SR	- 0.1444	0.0921	0.1214
FII	- 0.3143	0.1312	0.0193*

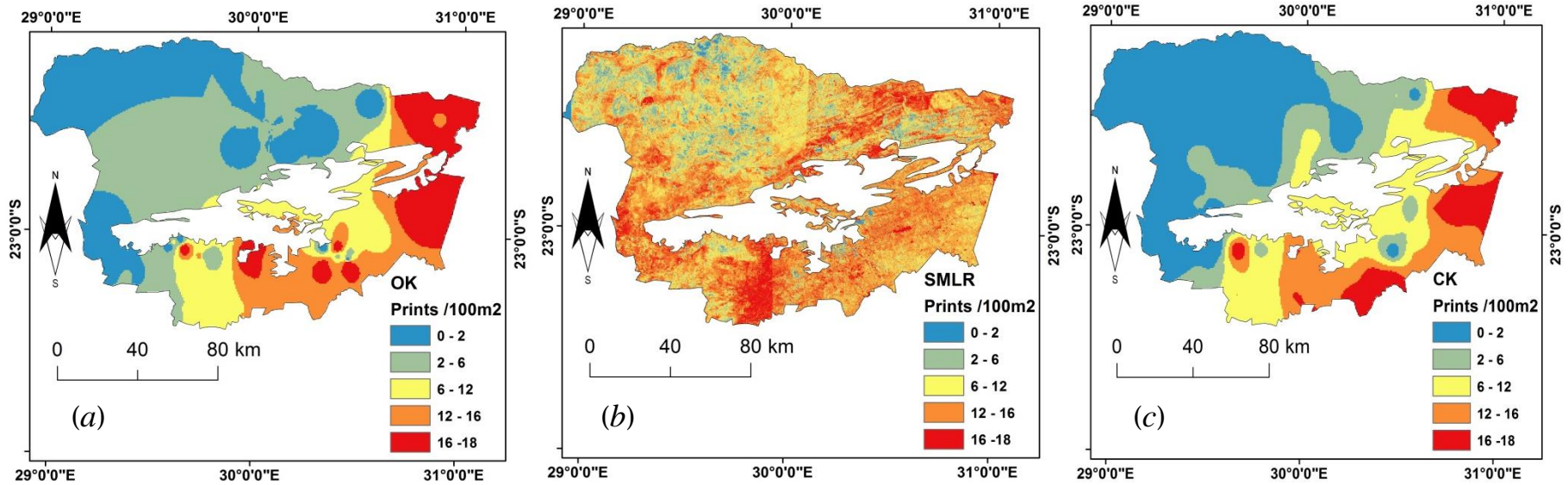
Note: significant levels:  $p < [*, 0.05]; [**, 0.01]; [***, 0.001]$ .

**Table 3.6:**

Comparison of interpolation performance among OK, CK and SMLR for predicting cattle hoofprints

Method	Validation		
	$R^2$	RMSE	MAD
OK	0.57	2.39	2.11
CK	0.69	0.20	0.04
SMLR	0.25	5.20	4.55

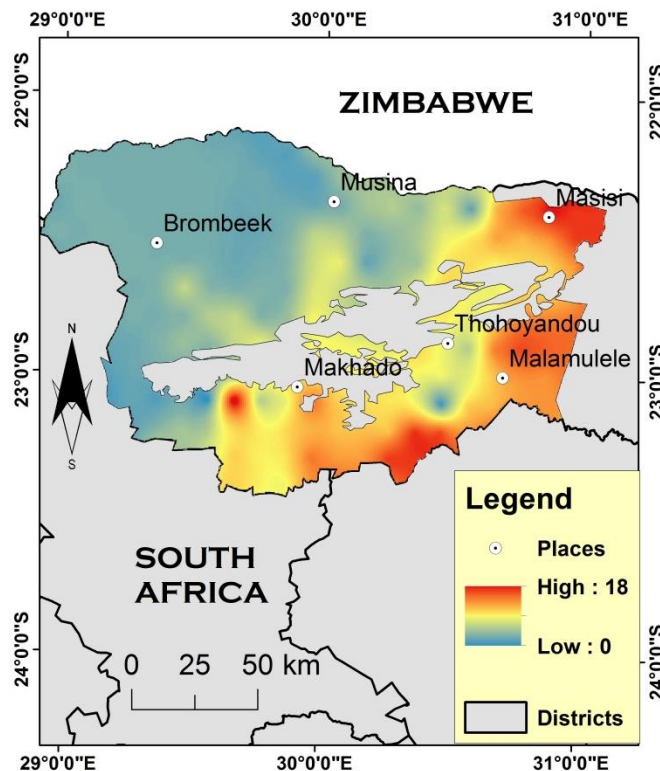
$R^2$  = coefficient of determination, RMSE = root mean square error, MAD = mean absolute deviance, OK = ordinary kriging, CK = co-kriging, SMLR = stepwise multiple linear regression.



**Figure 3.2:** Predicted cattle hoofprints values (per 100 m<sup>2</sup>) using (a) ordinary kriging, (b) step-wise multiple linear regression, and (c) co-kriging. The kriging methods were fitted using exponential model.

### 3.3.3. Mapping of cattle hoofprints

Figure 3.3 shows the final map produced from CK method used in the study. The trend of cattle hoofprints is that their number density is mainly high in the eastern and southern parts of the Vhembe District municipality. This map was produced from the variables that had the lowest RMSE, MAD and highest coefficient of correlation.



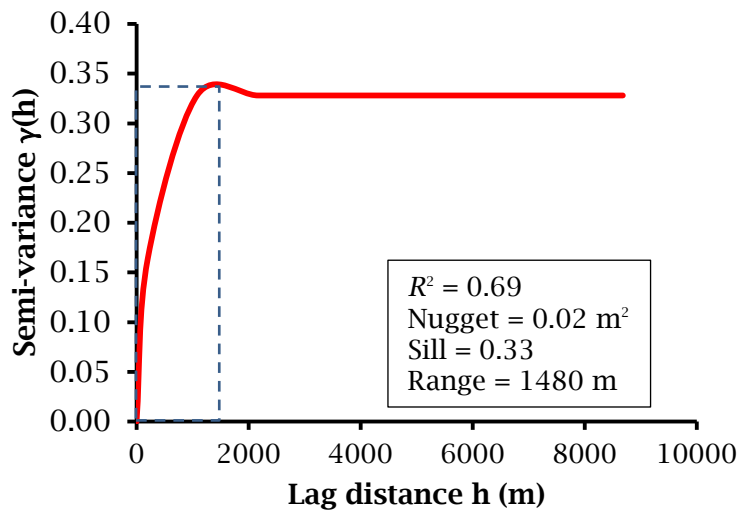
**Figure 3.3.** Predicted cattle hoofprints in the Vhembe District Municipality using co-kriging of field data and S-2 spectral data

### 3.4. Discussion

Cattle hoofprints have been documented to be one of the most crucial breeding habitats for *Anopheles* mosquito (Dutta *et al.*, 2010). However, although there exists a fair consensus on the importance of cattle hoofprints for malaria transmission, there has not been any attempt to date for mapping the distribution of such cattle hoofprints. Thus, in this chapter, we mapped the distribution of cattle hoofprints, by integrating field data with the S-2 imagery. The distribution of cattle in savannah grassland ecosystem has been reported elsewhere (Zengeya *et al.*, 2013; Kaszta *et al.*, 2017). However, no studies were done to estimate cattle hoofprint distribution by means of satellite derived data which are indicators of probable cattle hoofprint

occurrence. The results of this chapter are crucial for establishing possible links between malaria transmission and cattle hoofprint distribution, and for demonstrating the potential of satellite based mapping in malariometric studies (Adeola et al., 2017). They serve as the basis for targeted outdoor malaria control strategies since in VDM the use of indoor malaria control strategies is too far from being sufficient (Zhu et al., 2017). The interpolation techniques such as OK and CK have been effective in predicting the cattle hoofprints in unsampled areas. This was largely so due to the simplicity of method which relies on the neighbouring sample points to predict unmeasured ones (Yao et al., 2013). By combining S-2 dataset with field observation point data high prediction accuracy was achieved, which is indicate of the potential contribution of satellite data for characterizing *Anopheles* breeding micro-habitats. The findings in this chapter have shown that co-kriging interpolation method can be used for mapping micro-habitats of *An. arabiensis* and *An. fenestus* as long as the input data meet certain criteria. For example, it has been established that co-kriging can improve estimates of a less densely sampled primary variable (Wu et al., 2006). In instances where secondary variables contained skewed data, the data transformation is recommended (Saito and Goovaerts, 2000; Wu et al., 2006), as with the current study. This could have led to the higher hoofprint estimation in the current study, since co-kriging has high capability of utilizing cross-correlation factor between sparsely sampled primary (hoofprint point data) and densely sampled secondary ancillary variables (S-2 data) (Goovaerts, 1999; Adjorlolo and Mutanga, 2013). Additionally, the high  $R^2$  obtained though co-kriging may be attributed to kriging using more than one ancillary dataset. This is supported by findings from Wu et al. (2006) who demonstrated that ancillary data such as pH and organic carbon enhanced prediction ability of co-kriging method on plant-available zinc. This model characteristic gives the co-kriging an advantage over ordinary kriging and linear regression, which fail to account for cross-correlation between primary variable and multiple secondary variables.

Narrow bands of S-2 such as the new red edge1, 2 and 3 (694 nm – 908 nm) were significantly ( $p < 0.01$ ) correlated to the distribution of cattle hoofprints. These spectral bands, including SR and FII, have increased the co-kriging capability to predict cattle hoofprint distribution. By analysing the semivariogram of co-kriging method, one could realize that estimated range of spatial autocorrelation of cattle hoofprints was 1480 m (Figure 3.4).

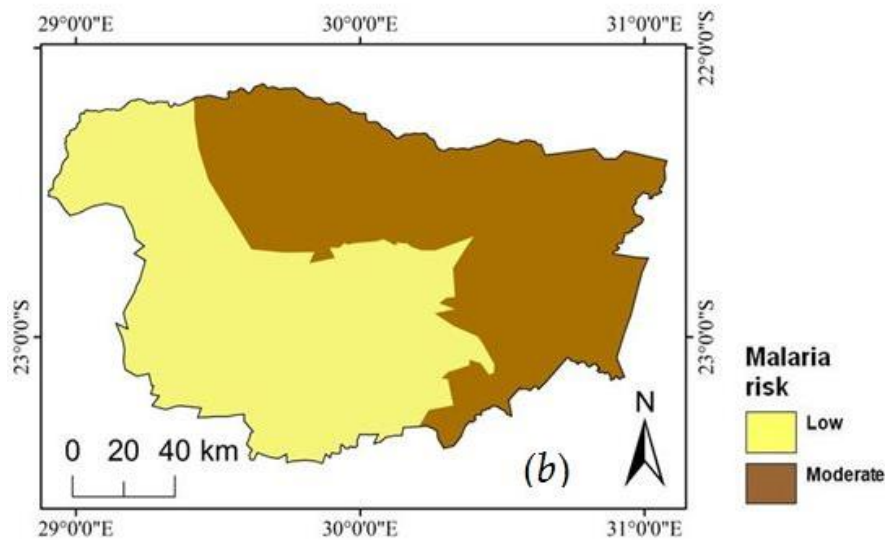


**Figure 3.4.** A semi-variogram for the co-kriging method using exponential model.

Correlations in the blue, red edge region, SRI suggest that the probabilities of finding cattle hoofprints increase when vegetation density and greenness decreases. Alternatively, as the area becomes more open the likelihood of finding cattle hoofprints increases. This is especially true considering that cattle are usually reared in environments that are more conspicuous, such as rangelands and open-field pastoral lands. The cattle livestock seldom graze in dense forests or thickets due to inaccessibility and the herdsmen's grazing practices (Zengeya *et al.*, 2013), although the associated grass is of high nutritive quality especially in dry season (Moyo *et al.*, 2013). Forests are mostly cooler than the areas with little to no vegetation, and as such are very hostile to *Anopheles* mosquito life cycle. The most ambient habitat for the productive *An. arabiensis*/*An. fenestus* larval population is the small puddles that are often surrounded by short grass (Ndenga *et al.*, 2011). In this chapter, most of the cattle hoofprints were estimated to have high likelihood of distribution in the north eastern, east and south central parts of the Vhembe District Municipality (Figure. 3.3). These areas have been documented to be areas of moderate to high malaria prevalence (Komen, 2017), although efforts are continuously made to decrease malaria transmission rates (Khosa *et al.*, 2013). Figure 3.5 shows the recently updated spatial distribution of malaria risk of the study area. From this map, it is evident that the areas of moderate malaria risk correspond to areas of high cattle hoofprint distribution, with an exception of the northern part of the study area. The prevalence of moderate malaria risk in the northern part of the study area may be attributed to the imported cases of malaria, largely by the patients from Zimbabwe (Raman *et al.* 2016). The villages within high cattle hoofprint zones are often characterized by very high rainfalls and periods of



high ambient temperatures (Komen *et al.*, 2014), which generally accelerate the vector life cycle and decrease pathogen incubation period of malaria parasite (Kovats and Martens, 2000).



**Figure 3.5.** The recently updated malaria risk map of at the study area. The malaria risk map was adapted from the map found at <http://traveldoctor.co.za/health/malaria/>.

However, in very dense grasslands it was very difficult to observe any cattle hoofprints, owing to the landscape vegetation composition and sampling protocol. Additionally, thick grass biomass create a layer on the ground which acts like a sponge when plodded, making it difficult for hoofprint to form. This chapter highlights the importance of both the conventional visible-near infrared band and the red edge indices although their in-depth contribution should further be tested. This study did not take into account multi-seasonal variations in vegetation phenology, landscape change and other surface prints which should be considered for future research. The study highlighted the capability of geostatistical application of remote sensing and field data for mapping cattle hoofprints, which are variable in nature depending on cattle rearing behaviour of the farmers (Kaszta *et al.*, 2017). Therefore, findings from this chapter offer a snapshot in time and their interpretations and implementations should be made with some level of caution. The effects of land use and land cover and the pastoral behaviour or preference by village herdsmen were not assessed and it is subject for future research.



### 3.5. Conclusions

The aim of this chapter was to explore the feasibility of S-2 and field method for mapping the distribution of cattle hoofprints in the Vhembe District Municipality. The following conclusions were drawn from this study:

- 1) High predictive accuracy was achieved through co-kriging of field observation footprint dataset and the S-2 data ( $R^2 = 0.69$ ). In addition the co-kriging method resulted in low predictive errors, with RMSE = 0.2 prints per 100 m<sup>2</sup>; MAD = 0.04 prints per 100 m<sup>2</sup>) using independent validation dataset.
- 2) S-2 data was correlated to the distribution of cattle hoofprints. The general trend was found to be significant correlations between blue, red edge (1-3) and Ferric Iron Index derived from S-2.
- 3) Most of the cattle hoofprints are found in the north-eastern (Masisi area), eastern (Malamulele/Makuleke area) and south central parts (Thohoyandou to Makhado areas) of the Vhembe District Municipality (Figure 3.3). These areas are known to comprise of high malaria incidence rates in the VDM, most of which are situated in the Mutale local municipality (Khosa *et al.*, 2013). Lower predictions were made for areas in the northern and western parts of the study area.

In summary, this chapter provides the baseline for mapping cattle hoofprints and further research is recommended in order to ascertain the findings. In this chapter cattle hoofprints (which are potential habitats for malaria vector) have been successfully mapped using a combination of remote sensing data and field data. This mapping was done to support the aim of the study and to address objective 1. Additionally, the architecture, composition of vegetation type and vegetation distribution affect the densities of mosquitoes that often quest and rest in the nearby trees around the homesteads. The following chapter presents the approach for mapping the vegetation leaf area index (LAI) using remote sensing data and the malaria dataset of the study area.

### 3.6. References

- Acharya B. (1999). Forest biodiversity assessment: a spatial analysis of tree species diversity in Nepal. PhD Thesis, ITC-Wageningen, the Netherlands.
- Adeola A.M., Botai, J.O., Rautenbach H., Adisa O.M., Ncongwane K.P., Botai C.M., Adebayo-Ojo T.C. (2017). Climatic Variables and Malaria Morbidity in Mutale Local Municipality, South Africa: A 19-Year Data Analysis. *International Journal of Environmental Research and Public Health*, 14(11):1360



- Adeola A.M., Olwoch J.M., Botai J.O., deW Rautenbach C.J., Kalumba A.M., Tsela P.L., Adisa O.M., Nsubuga F.W.N. (2015). Landsat satellite derived environmental metric for mapping mosquitoes breeding habitats in the Nkomazi municipality, Mpumalanga Province, South Africa. *South African Geographical Journal* 99(1): 14 – 28.
- Adjorlolo C., Mutanga O. (2013). Integrating remote sensing and geostatistics to estimate woody vegetation in an African savanna. *International Journal of Remote Sensing* 58(2): 305 – 322.
- Archibald S., Scholes R.J. (2007). Leaf green-up in semi-arid African savanna – separating tree and grass responses to environmental cues. *Journal of Vegetation Science* 18(4): 583 – 594.
- Barbet-Massin M., Jiguet F., Albert C.H., Thuiller W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2): 327 – 338.
- Birth G.S., Mcvey G.R. (1968). Measuring Color of Growing Turf with a Reflectance Spectrophotometer. *Agronomy Journal* 60: 640-649.
- Bouma M., Rowland M. (1995). Failure of passive zooprophyllaxis: cattle ownership in Pakistan is associated with a higher prevalence of malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 89: 351–353.
- Braack L., Hunt R., Koekemoer L.L., Gericke A., Munhenga G., Haddow A.D., Becker P., Okia M., Kimera I., Coetzee M. (2015). Biting behaviour of African malaria vectors: 1. where do the main vector species bite on the human body? *Parasites and Vectors*, 8(76): <https://doi.org/10.1186/s13071-015-0677-9>.
- Brooke B.D., Robertson L., Kaiser M.L., Raswiswi E., Munhenga G., Venter N., Wood O.R., Koekemoer L.L. (2015). Insecticide resistance in the malaria vector *Anopheles arabiensis* in Mamfene, KwaZulu-Natal. *South African Journal of Science* 111(11/12):1-3. <http://dx.doi.org/10.17159/sajs.2015/20150261>
- Burrough P.A., McDonnell R.A. (1998). Principles of Geographical Information Systems. Spatial Information Systems and Geostatistics. Oxford University Press, Oxford.
- Carter G.P., Miskewits R.J., Isukapalli S., Mun Y., Vyas V., Yoon S., Georgeopoulos P., Uchrin C.G. (2011). Comparison of kriging and cokriging for the geostatistical estimation of specific capacity in the Newark Basin (NJ) aquifer system. *Journal of Environmental Science and Health: Part A, Toxic/ Hazardous Substances Environmental Engineering*, 46(4): 371 – 377.



- Charlwood J.D., Macia G.A., Manhaca M., Sousa B., Cuamba N., Bragança M. (2013). Population dynamics and spatial structure of human-biting mosquitoes, inside and outside of houses, in the Chockwe irrigation scheme, southern Mozambique. *Geospatial Health* 7(2): 309 – 320.
- Chen D., Huang J., Jackson T.J. (2005). Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands. *Remote Sensing of Environment* 98: 225–236.
- Clark R.N. (1999). Spectroscopy of rocks and minerals, and principles of spectroscopy, in Rencz, A.N., ed., *Remote sensing for the earth sciences*, in Ryerson, R.A., ed., *Manual of remote sensing*, v. 3: New York, John Wiley & Sons, Inc., p. 3–58.
- Clennon J.A., Kamanga A., Musapa M., Shiff C., Glass G.E. (2010). Identifying malaria vector breeding habitats with remote sensing data and terrain-based landscape indices in Zambia. *International Journal of Health Geographics*, 9(58): <https://doi.org/10.1186/1476-072X-9-58>.
- Cohen W.B., Maieresperger T.K., Gower S.T., Turner D.P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, 84(4): 561 – 571.
- Davies C. (2016). Influence of environmental characteristics on the habitat of and behavioural interactions between *Anopheles* species in South Africa. *Master of Science Thesis, University of Witwatersrand, South Africa*.
- Deiss L., Franzluebbers A.J., de Moraes A (2017). Soil texture and organic carbon fractions predicted from near-infrared spectroscopy and geostatistics. *Soil Science Society of American Journal*, 81(5):1222-1234.
- Delegido J., Verrelst J., Meza C.M., Rivera J.P., Alonso L., Moreno J. (2013). A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *European Journal of Agronomy* 46: 42–52.
- Dlamini S.N., Franke J., 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(1): DOI: 10.4081/gh.2015.302.
- Dutta P., Khan S.A., Bhattacharyya D.R., Khan A.M., Sharma C.K., Mahanta J. (2010). Studies on the breeding habitats of the vector mosquito *Anopheles baimai* and its relationship to malaria incidence in Northeastern region of India. *EcoHealth* 7(4): 498 – 506.

- Emidi B., Kisinza W.N., Mmbando B.P., Malima R., Mosha F.W. (2017). Effect of physicochemical parameters on *Anopheles* and *Culex* mosquito larvae abundance in different breeding sites in a rural setting of Muheza, Tanzania. *Parasites and Vectors* 10(304): <https://doi.org/10.1186/s13071-017-2238-x>.
- Exelis Visual Information Solutions. (2017). Environment for Visualizing Images. Boulder, CO: Exelis Visual Information Solutions.
- Fotheringham A.S., Brunson C., Charlton M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Chichester: Wiley.
- Franco A.O., Davies C.R., Coleman P.G. Effects of Livestock Ownership on the Risk of Human Malaria: a Case Control Study in Ethiopia. *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006, Cairns, Australia*. Franco A.O., Gabriela M., Gomes M., Rowland M., Coleman P.G., Davies C.R. (2014). Controlling Malaria Using Livestock-Based Interventions: A One Health Approach. *PLoS ONE*, 9(7): e101699. doi.org/10.1371/journal.pone.0101699.
- Hmimina G, Dufrêne E., Pontailleur J.Y., Delpierre N., Aubinet M., Caquet B., de Grandcourt A., Burban B., Flechard C., Granier A., Gross P., Heinesch B., Longdoz B., Moureaux C., Ourcival J.M., Rambal S., Saint André L., Soudani K. (2011). Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements. *Remote Sensing of Environment*, 132: 145-158.
- Gao B.C. (1996). NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58: 257–266.
- Gao H., Wang L., Jing L., Xu J. (2016). An effective modified water extraction method for Landsat-8 OLI imagery of mountainous plateau regions. *9th Symposium of the International Society for Digital Earth (ISDE); IOP Conference Series: Earth and Environmental Science*, 34:012010; doi:10.1088/1755-1315/34/1/012010.
- Gerritsen A.A.M., Kruger P., Schim van der Loeff M.F., Grobusch M.P. (2008). Malaria incidence in Limpopo Province, South Africa, 1998 – 2007. *Malaria Journal* 7:1 – 8.
- Gimnig J.E., Ombok M., Otieno S., Kaufman M.G., Vulule J.M., Walker E.D. (2002). Density-dependant development of *Anopheles gambiae* (Diptera: Culicidae) larvae in artificial habitats. *Journal of Medical Entomology* 39: 162 – 172.
- Gitelson A., Merzlyak M.N. (1994). Spectral reflectance changes associated 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.

- Goovaerts P. (1999). Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89:1 – 45.
- Habtewold T. (2004) Interaction between *Anopheles*, cattle and human: exploration of the effects of various cattle management practices on the behaviour and control of *Anopheles arabiensis* in Ethiopia. PhD Thesis. Greenwich: University of Greenwich, U.K. Page: 249.
- Highton R.B., Bryan J.H., Boreham P.F.L., Chandler J.A. (1979). Studies on sibling species *Anopheles gambiae* Giles and *Anopheles arabiensis* Patton (Diptera: Culicidae) in the Kisumu area, Kenya. *Bulletin of Entomological Research*, 69: 43-53.
- Huang J., Walker E.D., Otienoburu P.E., Amimo F., Vulule J., Miller J.R. (2006). Laboratory tests of oviposition by the African malaria mosquito, *Anopheles gambiae*, on dark soil as influenced by presence or absence of vegetation. *Malaria Journal*, 5:88. DOI: <https://doi.org/10.1186/1475-2875-5-88>
- Hurd H. “Can cows protect against bites?” BugBitten (BMC). 27 Mar 2014. <http://blogs.biomedcentral.com/bugbitten/2014/03/27/can-cows-protect-against-mosquito-bites-2/>
- Isaaks E.H., Srivastava R.M. (1989). An Introduction to Applied Geostatistics, Oxford University Press, 561 pages
- Iwashita H., Dida G.O., Sonye G.O., Sunahara T., Futami K., Njenga S., Chaves L.F., Minakawa N. (2014). Push by a net, pull by a cow: can zooprophylaxis enhance the impact of insecticide treated bed nets on malaria control? *Parasites and Vectors*, 7(52): <https://doi.org/10.1186/1756-3305-7-52>.
- Jackson R.D., Huete A.R. (1991). Interpreting vegetation indices. *Preventative Veterinary Medicine*, 11(3-4): 185 – 200.
- Jacob B.G., Muturi E.J., Mwangangi J.M., Funes J., Caamano E.X., Muriu S., Shililu J., Githure J., Novak R.J. (2007). Remote and field level quantification of vegetation covariates for malaria mapping in three rice agro-village complexes in Central Kenya. *International Journal of Health Geographics*, 6(21): DOI: 10.1186/1476-072X-6-21.
- Jordan C.F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 50: 663–666.
- Joshi G.P., Service M.W., Pradhan G.D. (1975). A survey of species A and B of the *Anopheles gambiae* Giles complex in the Kisumu area of Kenya prior to insecticidal

spraying with OMS-43 (Fenitrothion). *Annals of Tropical Medicine and Parasitology*, 69: 91-104.

- Kambhammettu B.V.N.P., Allena P., King J.P. (2011). Application and evaluation of universal kriging for optimal contouring of groundwater levels. *Journal of Earth Systems Science* , 120(3): 413 – 422.
- Kaszta Z., Marino J., Wolff E. (2017). Fine-scale spatial and seasonal rangeland use by cattle in a foot-and-mouth disease control zones. *Agriculture, Ecosystems and Environment* 239: 161–172.
- Khosa E., Kuonza L.R., Kruger P., Maimela E. (2013). Towards the elimination of malaria in South Africa: a review of surveillance data in Mutale Municipality, Limpopo Province, 2005 to 2010. *Malaria Journal*, 12(1):7.
- Kokaly R.F., Clark R.N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing Environment*, 67:267–287.
- Komen K. (2017). Could malaria control programmes be timed to coincide with the onset of rainfall? *EcoHealth* . DOI: 10.1007/s10393-017-1230-4.
- Komen K., Olwoch J., Rautenbach H., Botai J., Adebayo A. (2014). Long-Run Relative Importance of Temperature as the Main Driver to Malaria Transmission in Limpopo Province, South Africa: A Simple Econometric Approach. *EcoHealth* DOI: 10.1007/s10393-014-0992-1.
- Kovats R.S., Martens P. (2000). Human health. *In: Assessment of Potential Effects and Adaptations for Climate Change in Europe: The Europe ACACIA Project*. Parry ML (editor), Norwich: Jackson Environment Institute, University of East Anglia, pp 227–242.
- Lacerda M.P.C., Demattê J.A.M., Sato M.V., Fongaro C.T., Gallo B.C., Souza A.B. (2016). Tropical Texture Determination by Proximal Sensing Using a Regional Spectral Library and Its Relationship with Soil Classification. *Remote Sensing* 8(701): doi:10.3390/rs8090701.
- Lobo N. (2010). *Malaria in the Social Context: A Study in Western India*, Routledge, 2 Park Square, Milton park, Abingdon, Oxon.
- Mahande A., Mosha F., Mahande J., Kweka E. (2007). Feeding and resting behaviour of malaria vector, *Anopheles arabiensis* with reference to zooprophyllaxis. *Malaria Journal* , 6(100): <https://doi.org/10.1186/1475-2875-6-100>.





- Malahlela O.E. (2016). Inland waterbody mapping: towards improving discrimination and extraction of inland surface water features. *International Journal of Remote Sensing*, 37(19): 4574 – 4589.
- Malahlela O.E., Cho M.A., Mutanga O. (2014). Mapping canopy gaps in an indigenous subtropical coastal forest using high resolution WorldView-2 data. *International Journal of Remote Sensing* 35(17): 6397 – 6417.
- Malahlela O.E., Olwoch J.M., Adjorlolo C. (2018). Evaluating Efficacy of Landsat-Derived Environmental Covariates for Predicting Malaria Distribution in Rural Villages of Vhembe District, South Africa. *EcoHealth* <https://doi.org/10.1007/s10393-017-1307-0>
- Marchetti A., Piccini C., Francaviglia R., Mabit L. (2012). Spatial distribution of soil organic matter using geostatistics: A key indicator to assess soil degradation status in central Italy. *Pedosphere* 22(2): 230–242.
- Martinez B., Cassiraga E., Camacho F., Garcia-Haro J. (2010). Geostatistics for Mapping Leaf Area Index over a Cropland Landscape: Efficiency Sampling Assessment. *Remote Sensing* 2(11):2584 – 2606.
- Mayagaya V.S., Nkwengulila G., Lyimo I.N., Kihonda J., Mtambala H., Ngonyani T., Russell T.L., Furguson H.M. (2015). The impact of livestock on the abundance, resting behaviour and sporozoite rate of malaria vectors in southern Tanzania. *Malaria Journal*, 14(17): <https://doi.org/10.1186/s12936-014-0536-8>.
- McLaughlin R.E., Vidrine M.F. (1987). *Psorophora columbiae* larval density in SouthWestern Louisiana rice fields as a function of cattle density. *Operational and Scientific Notes*, 3(4): 633 – 635.
- Moleele N.M., Perkins J.S. (1998). Encroaching woody plant species and boreholes: is cattle density the main driving factor in the Olifants Drift communal grazing lands, south-eastern Botswana. *Journal of Arid Environments* 40(3): 245 – 253.
- Moonasar D., Asomugha C., Baker L., Blumberg L., Barnes K.I., Maharaj R., Benson F. (2011). Preventing disease and saving lives: the malaria season is upon us. *South African Medical Journal* 101(12): 865. doi:10.7196/SAMJ.5345.
- Moyo B., Dube S., Lesoli M., Masika P., (2013). Seasonal habitat use and movement patterns of cattle grazing different rangeland types in the communal areas of the Eastern Cape, South Africa. *African Journal of Agricultural Research* 8: 36–45. doi:http://dx.doi.org/10.5897/AJAR12.222.
- Müller-Wilm U. (2016). S-2 MSI – Level-2A Prototype Processor Installation and User Manual. *S2PAD-VEGA-SUM-0001*, Issue (2.2).



- Munga S., Yakob L., Mushinzimana E., Zhou G., Ouna T., Minakawa N., Githeko A., Yan G. (2009). Land Use and Land Cover Changes and Spatiotemporal Dynamics of *Anophele* Larval Habitats during a Four-Year Period in a Highland Community of Africa. *The American Journal of Tropical Medicine and Hygiene*, 81(6): 1079 – 1084.
- Mutanga O., Rugege D. (2006). Integrating remote sensing and spatial statistics to model herbaceous biomass distribution in a tropical savanna. *International Journal of Remote Sensing*, 27, 3499-3514.
- Mutanga O., Skidmore A.K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19): 3999 – 4014.
- Myers V.I., Heilman M.D., Lyon R.J.P., Namken L.N., Simonett D., Thomas J.R., Wiegand C.L., Woolley J.T. (1970). Soil, water and plant relations. In *Remote Sensing with Special Reference to Agriculture and Forestry*, pp. 253 – 297 (Washington, DC: National Academy of Sciences).
- Ndenga B.A., Simbauni J.A., Mbugi J.P., Githeko A.K., Fillinger U. (2011). Productivity of malaria vectors from different habitat types in the western Kenya highlands. *PLoS ONE* ;6:e19473.
- Omran ESE (2012) Detection of land-use and surface temperature change at different resolutions. *Journal of Geographic Information Systems* 4:189–203.doi:10.4236/jgis.2012.43024.
- Pfaehler O., Oulo D.O., Gougna L.C., Githure J., Guerin P.M. (2006). Influence of soil quality in the larval habitat on development of *Anopheles gambiae* Giles. *Journal of Vector Ecology*, 31(2):400 – 405.
- Pietola L., Horn R., Yli-Halla M. (2005). Effects of trampling by cattle on the hydraulic and mechanical properties of soil. *Soil and Tillage Research*, 82(1): 99 – 108.
- Ramudzuli M.R., Horn A.C. (2014). Arsenic residues in soil at cattle dip tanks in the Vhembe district, Limpopo Province, South Africa. *South African Journal of Science* 110(7/8).Art#2013-0393: 1 – 7.
- Reza S., Nayak D., Mukhopadhyay S., Chattopadhyay T., Singh S. (2017). Characterizing spatial variability of soil properties in alluvial soils of India using geostatistics and geographical information system. *Archives of Agronomy and Soil Science*, 63(11):1489-1498.
- Robinson T.P., Metternicht G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture* 50(2): 97 – 108.



- Rouse J.W., Haas R.H., Schell J.A. (1974). Monitoring the Vernal Advancement and Retrogradation (Greenwave Effect) of Natural Vegetation. Texas A and M University, College Station.
- Saito H., Goovaerts P. (2000). Geostatistical interpolation of positively skewed and censored data in a dioxin-contaminated site. *Environmental Sciences and Technology*, 34: 4228 – 4235.
- Sande S., Zimba M., Chinwada P., Masendu H.T., Makuwaza A. (2015). Malaria vector species composition and the relative abundance in the Mutare and Mutasa districts, Zimbabwe. *Journal of Entomological and Acarological Research*, 47(4955): 79 – 85.
- Sattler M.A., Mtasiwa D., Kiama M., Premji Z., Tanner M., Killeen G.F., Lengeler C. (2005). Habitat characterization and spatial distribution of *Anopheles* sp. mosquito larvae in Dar es Salaam (Tanzania) during an extended dry period. *Malar J.* 2005, 4(1):4.
- Setianto A., Triandini T. (2013). Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. *Journal of South-East Asian Applied Geology*, 5(1): 21 – 29.
- Sewe M.O., Tozan Y., Rocklov J. (2017). Using remote sensing environmental data to forecast malaria incidence at a rural district hospital in Western Kenya. *Scientific Reports*, 7: 2589. DOI: s41598-017-02560-z.
- Shit P.K., Bhunia G.S., Maiti R. (2016). Spatial analysis of soil properties using GIS based geostatistical models. *Modelling Earth Systems and Environment* 2(107): <https://doi.org/10.1007/s40808-016-0160-4>
- Shit P.K., Bhunia G.S., Maiti R. (2016). Spatial analysis of soil properties using GIS based geostatistics models. *Modelling Earth Systems and Environment*, 2(2):107.
- Sinka M.E., Bangs M.J., Manguin S., Coetzee M., Mbogo C.M., Hemingway J., Patil A.P., Temperley W.H., Gething P.W., Kabaria C.W., Okara R.M., van Boeckel T., Godfray H.C.J., Harbach R.E., Hay S.I.(2010). The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic precis. *Parasite and Vectors*, 3: 117-10.1186/1756-3305-3-117.
- Song G., Zhang J., Wang K. (2014). Selection of optimum auxiliary soil nutrient variables for co-kriging interpolation. *PLoS ONE*, 9(6): e99695.
- Tabachnick, B.G., Fidell, L.S., 1996. Using Multivariate Statistics. Harper Collins, New York.

- Ustin, S. L., A. A. Gitelson, S. Jacquemoud, M. Schaepman, G. P. Asner, J. A. Gamon, and P. Zarco-Tejada. 2009. "Retrieval of Foliar Information about Plant Pigment Systems from High Resolution Spectroscopy." *Remote Sensing of Environment* 113 (Suppl. 1): S67–S77. doi:10.1016/j.rse.2008.10.019.
- Vincini M., Frazzi E., D'Alessio P. (2008). A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precision Agriculture* 9: 303–319.
- W.H.O. (1982). Manual on environmental management for mosquito control with special emphasis on mosquito vectors. Offset Publication No. 66. Geneva: World Health Organization. Page: 283.
- Webster R., Oliver M.A. (2001). *Geostatistics for Environmental Scientists* John Wiley and Sons, Brisbane, Australia.
- Wessels K.J., Prince S.D., Frost P.E., Van Zyl D. (2004). Assessing the effects of Human-Induced Land Degradation in the Former Homelands of Northern South Africa with a 1km AVHRR NDVI Time Series. *Remote Sensing of Environment* 91:47-67.
- World Health Organization (WHO). (2015a). Malaria: draft global technical strategy: post 2015. *Sixty-eighth World Health Assembly, Provisional agenda item*, 16.2 (A68/28).
- World Health Organization (WHO). (2015b). World Malaria Report.
- Wu J., Norvell W.A., Welch R.M. (2006). Kriging on highly skewed data for DTPA-extractable soil Zn with auxiliary information for pH and organic carbon. *Geoderma*, 134: 187 – 199.
- Yao X., Fu B., Lü Y., Sun F., Wang S., Liu M. (2013). Comparison of Four Spatial Interpolation Methods for Estimating Soil Moisture in a Complex Terrain Catchment. *PLoS ONE*, 8(1): e54660. doi:10.1371/journal.pone.0054660.
- Zengeya F.M., Mutanga O., Murwira A. (2013). Linking remotely sensed forage quality estimates from WorldView-2 multispectral data with cattle distribution in a savanna landscape. *International Journal of Applied Earth Observation and Geoinformation*, 21:513–524.
- Zhou et al., Huang F., Wang J.J., Zhang S.S., Su Y.P., Tang L.H. (2010). Geographical, meteorological and vectorial factors related to malaria re-emergence in Huang-Huai River of central China. *Malaria Journal*, 9(337): <https://doi.org/10.1186/1475-2875-9-337>.
- Zhu L., Müller G.C., Marshall J.M., Arheart K.L., Qualls W.A., Hlaing W.M., Schlein Y., Traore S.F., Dombia S., Beier J.C.(2017). Is outdoor vector control needed for malaria



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

elimination? An individual-based modelling study. *Malaria Journal*, 16: 266. doi:  
10.1186/s12936-017-1920-y.

- Zha, Y., Gao J., Ni S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* 24(3):583–594.
- Zhou G., Minakawa N., Githeko A.K., Yan G. (2004). Association between climate variability and malaria epidemics in the East African highlands. *Proceedings of the National Academy of Sciences of the United States of America* 101(8): 2375 – 2380.
- Zhou S., Zhang S., Wang J., Zheng X., Huang F., Li W., Zhang H. (2012). Spatial correlation between malaria cases and water-bodies in *Anopheles sinensis* dominated areas of Huang-Huai plain, China. *Parasites and Vectors* 5(106):1 – 7.



## CHAPTER 4<sup>3</sup>:

# Comparison of the Sentinel-2 broadband and narrowband vegetation indices for mapping LAI in malaria-prone heterogeneous semi-arid environment of Southern Africa

Oupa E. Malahlela <sup>a,b\*</sup>, Adjorlolo C <sup>c.</sup>, Olwoch J.M. <sup>a,d</sup>

<sup>a</sup> *Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa*

<sup>b</sup> *South African National Space Agency (SANSA), Earth Observation Directorate, Pretoria 0001, South Africa*

<sup>c</sup> *New Partnership for Africa's Development (NEPAD) Agency, 230 15<sup>th</sup> Road, Midrand, South Africa*

<sup>d</sup> *Southern African Science Service Center for Climate Change and Adaptive Land Management (SASSCAL), Windhoek 91100, Namibia*

***This paper relates to objectives 2 of the thesis.***

---

<sup>3</sup>This chapter is based on the manuscript titled “*Comparison of the Sentinel-2 broadband and narrowband vegetation indices for mapping LAI in malaria-prone heterogeneous semi-arid environment of Southern Africa*” (GIScience and Remote Sensing – in review).



## Abstract

Leaf area index (LAI) is an important parameter defining ecosystem services, functioning, carbon budgets, and micro-habitat temperature regulations for poikilothermic organisms. Spatial mapping of LAI variation will essentially contribute to the understanding of how vegetation structural parameters affect malaria vector questing, resting and eventual spread of infectious diseases in malaria prone environment. In this chapter, we compared the performance of Sentinel-2 derived broad-band vegetation indices (BBVI) and the narrow-band vegetation indices (NBVI) for mapping the distribution of LAI in a heterogeneous landscape of the Vhembe District Municipality, South Africa. The Sentinel-2 bands were resampled to 10 m spatial resolution to allow for comparison. The results from this chapter indicate that the BBVI called modified chlorophyll absorption ratio index (MCAR<sub>l2</sub>) and the modified triangular vegetation index (MTVI<sub>l2</sub>) have shown higher correlation with the LAI distribution ( $R^2 = 0.73$ ; RMSE = 0.86 m<sup>2</sup> m<sup>-2</sup>) than the NBVI computed from the red edge region ( $R^2 = 0.61$ ; RMSE = 1.61 m<sup>2</sup> m<sup>-2</sup>). This chapter emphasizes the importance of the Sentinel-2 broad bands (visible to near-infrared), acquired at 10 m spatial resolution which is the important finding in this study. Findings from this chapter indicate a robust generalization of LAI estimation, because the parameterization was made from landscape with heterogeneous vegetation communities. The high resolution LAI map derived from Sentinel-2 could form part of the long-term strategy for environmental monitoring in a heterogeneous environment.

**Keywords:** Leaf area index; high resolution; Sentinel-2; heterogeneous vegetation

---

## 4.1. Introduction

Owing to the increase in the use of intradomiciliary-based control measures such as indoor residual spraying (IRS) and long-lasting insecticide-treated nets (LLINs), there has been substantial decline in malaria transmission in Southern Africa (Mayagaya *et al.*, 2015). However, malaria transmission still persists in many parts of Southern Africa, owing to persistent socio-economic, outdoor environmental factors and climatic change influencing malaria vector ecology. In recent years, extensive efforts to control *Anopheles* mosquitoes have focused on development of outdoor strategies targeting vector larvae (Dewald *et al.*, 2016). Such efforts have targeted malaria vectors such as *An. arabiensis* whose transmission events are predominantly estimated to occur outdoors (Killen *et al.*, 2016). Accurate knowledge on the vector biology and the outdoor environment could enhance the efficiency of malaria control strategies and thus reducing new cases of malaria transmission.



As small poikilothermic organisms, mosquitoes ([Afrane et al., 2004](#)) require shaded areas for the most of the day in order to avoid severe desiccation and heat stress that may result from direct sunlight in tropical location ([Dewald et al., 2016](#); [Paaijmans and Thomas, 2011](#)). Previous studies have shown that vegetation provides mosquitoes with resting and refuge sites. ([Ricotta et al., 2014](#)), thus facilitating malaria risk to nearby settlements. Whether the adult mosquito prefers one vegetation type as a refuge against possible predators (e.g. dragonfly) over another will depend on the architecture of particular plant stand. Individual plant stands have a dominant role in maintaining malaria transmission by defining gradients of light, moisture and temperature ([Clark et al., 2005](#)) among other factors. However, these ecological functions depend largely on the biophysical characteristics of plant and edaphic factors surrounding it ([Moro et al., 2015](#); [Clark et al., 2015](#)).

The leaf area index (LAI) is one of the fundamental bio-indicators of plant physiologic conditions ([Sampson et al., 2003](#)), and plant functioning ([Houborg et al., 2015](#)). LAI is geometrically defined as the total one-sided leaf area per ground surface area ([Breda 2003](#); [Darvishzadeh et al., 2008](#)), and is usually expressed in terms of  $m^2 m^{-2}$  ([Asano et al., 2009](#)). Over the years, remote sensing has been recognized as the effective and reliable method for estimating vegetation biophysical characteristics such as LAI ([Houborg and Boegh, 2008](#); [Carmona et al., 2015](#); [Clevers et al., 2017](#)). However, most of the remote sensing approaches for retrieving LAI focused primarily on agricultural crop monitoring and the homogenous vegetation communities ([Viña et al., 2011](#); [Masemola et al., 2016](#);). Moreover, the parameterization of vegetation LAI is commonly derived for homogeneous agricultural crops due to its significance and robustness in precision agriculture ([Haboudane et al., 2002](#); [Hunt Jr. et al., 2013](#)). However, the robustness of LAI mapping for heterogeneous vegetation communities has received less attention, and thus it remains to be seen how robust is the LAI calibrated from heterogeneous landscape.

There is a limited number of studies that implemented LAI retrieval methods on heterogeneous landscape, especially in savannah environment ([Tong and He, 2017](#); [Darvishzadeh et al., 2008b](#)). It is crucial, however, to assess the application of remote sensing for retrieving LAI at heterogeneous savannah landscape in order to understand spatial variations of plant physiology and photosynthetic capabilities, and how the vegetation LAI impacts on malaria prevalence. This is a fundamental aspect when defining vegetation function and composition that may be associated with habitats of infectious diseases vectors such as *Anopheles arabiensis* and *Anopheles fenestus* mosquitoes. Studies in optical remote sensing have

shown that the LAI is correlated with specific regions of the vegetation spectra (Gausman, 1982) and such correlation may be retrieved through various methods. A number of approaches have been used for retrieving the LAI from the remote sensing datasets. Amongst these approaches, two methods are common with the current remote sensing developments: (i) inversion of canopy radiative transfer models (RTMs), such as PROSAIL model (Jadquemoud *et al.*, 2009; Masemola *et al.*, 2016) and (ii) the empirical relationship between LAI and the spectral vegetation indices (Glenn *et al.*, 2008; Xavier and Vettorazzi, 2004). The RTMs have the advantage of explicitly accounting for all factors that influence canopy reflectance and architecture. However, the RTMs often have ill-posed problem associated with similar remote sensing signal due to combination of canopy biophysical and biochemical variables that have mutually compensating effect on canopy reflectance (Dorigo *et al.*, 2009). On the other hand, parameterization and regularization of RTMs often render them complicated and may not account for optimized variability caused by the larger spatial coverage of the biophysical variables (Din *et al.*, 2017; Ryu *et al.*, 2009). The vegetation indices (VIs) have been successfully correlated to structural characteristics of plants, thus contributing significantly to the more effective alternative approach for retrieving LAI (Xie *et al.*, 2014). The use of VIs is more widespread due to the ease of computation and interpretation. The VIs in remote sensing are the semi-analytical mathematically transformed indices from multiple spectral bands ( $R_n$ ) with the aim of minimizing the variability of external factors such as soil background, so as to enhance the vegetation biophysical/chemical characteristics (Darvishzadeh *et al.*, 2006).

Various forms of VIs were explored to assess the empirical relationships between spectral reflectance data and LAI (Borzuchowski and Schulz, 2010). Perhaps one of the most commonly used VIs is the normalized difference vegetation index (NDVI), which is computed from the reflectance of the red and near infrared spectral bands. As a broad-band vegetation index (BBVI), the NDVI has been used for retrieval of LAI for many years (Breunig *et al.*, 2011; Sun *et al.*, 2017). However, the non-linearity of the relationship between NDVI and LAI suffers a rapid decrease of sensitivity particularly in vegetation of moderate-to-high densities of photosynthetic green biomass (Gitelson *et al.*, 2007). Alternative methods have been proposed that yield more linear relationship between the LAI and remote sensing data. A study by Haboudane *et al.* (2004) has demonstrated that the modified chlorophyll absorption ratio index (MCARI<sub>2</sub>) and the modified triangular vegetation index (MTVI<sub>2</sub>) were the best predictors of green LAI levels in croplands. On the other hand, some studies have shown that the performance of narrow-band vegetation indices (NBVI) surpasses that of BBVI for estimating



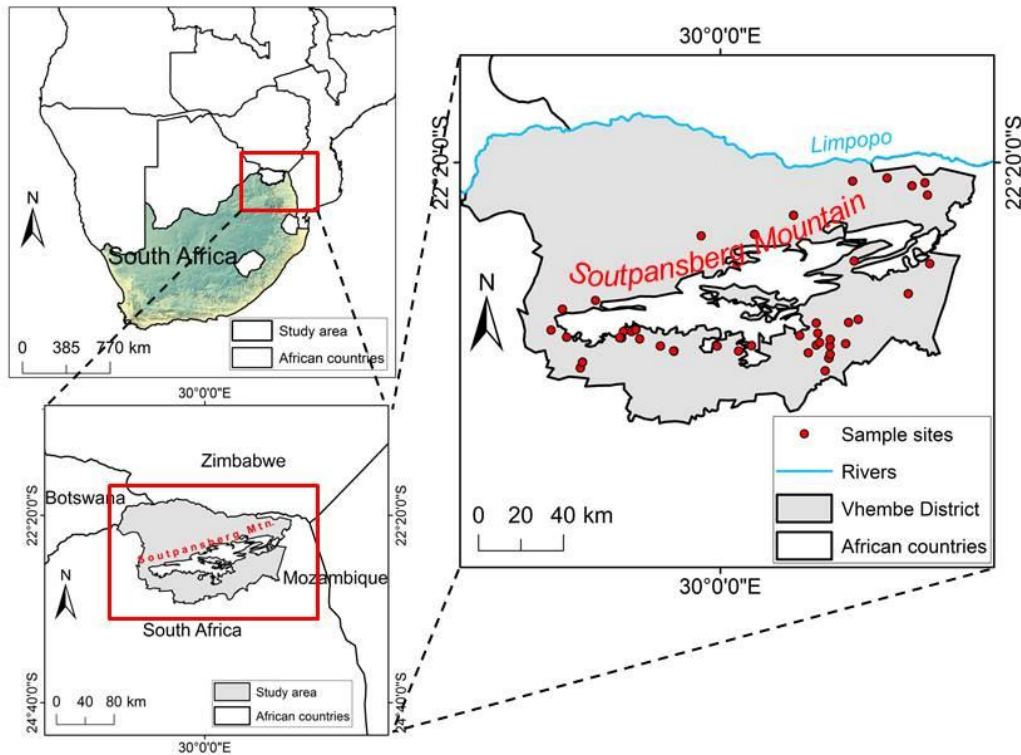
vegetation biophysical and biochemical characteristics because, due to the sensitivity of NBVI to subtle vegetation physiochemical parameters (Mutanga and Skidmore, 2004). Granted, there seems to be disagreements of results from various studies concerning the advantage of NBVI over the conventional BBVI, or vice versa, for estimating LAI. For example, more recently Clevers *et al.* (2017) tested the application of Sentinel-2 BBVI and NBVI for mapping leaf chlorophyll content, canopy chlorophyll content and the LAI of potato crops, and have concluded that the BBVI exhibited high mapping accuracy than the NBVI, thereby avoiding the need for narrow-bands that are available at 20 m spatial resolution.

This chapter aimed at comparing the BBVI and NBVI of Sentinel-2 for mapping the green LAI in heterogeneous landscape of the Vhembe district in South Africa, which is mostly comprised of natural vegetation. Sentinel-2 was a preferred sensor because it comprises of both broadband visible-near infrared bands (10 m) and four narrow bands in the red edge region (20 m). In addition, Sentinel-2 has a repeat cycle of 5 days, which is one of the key properties for monitoring potential *Anopheles* mosquito resting site, in the form of LAI.

## 4.2. Methods

### 4.2.1. Study area

The study was conducted in the Vhembe District Municipality (VDM), which is located in the Limpopo province of South Africa. The study area covers a land area of approximately 25 596 km<sup>2</sup>. Detailed description of the study area is found in Malahlela *et al.*, 2018. (Figure 4.1) While in predominantly savannah biome, the VDM comprises of a mixture of seven (7) veld types, viz. Mopani veld (largest), Lowveld sour bushveld, Mixed bushveld, arid sweet bushveld, North-eastern sour veld, sourish mixed bushveld, and sour bushveld (smallest) (LSER, 2004). The density of vegetation in the study area ranges from very sparse vegetation in the west to very dense deciduous vegetation in the east and south central parts of the study area. One of the most characteristic vegetation species is the *Adansonia digitata* (baobab tree) commonly found in this area (Rutherford *et al.*, 2006). Both the *Acacia* (*spp*) and *Dichrostachys cinerea* were found to be the dominant shrub species in the south-western part of the study area, while *D. cinerea* dominates the eastern part of the study area (researchers' observation). The study area usually receives summer rainfalls, which ranges from 200 mm to over 1000 mm annually (LSER, 2004), although there are frequent occurrences of flash floods in the area (Reason and Keibel, 2004).



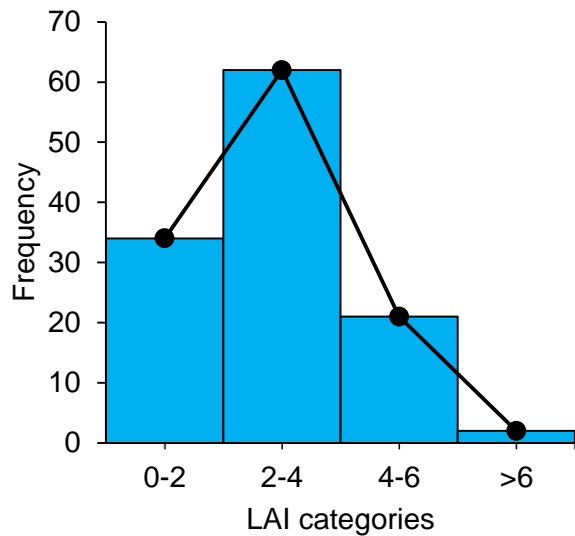
**Figure 4.1.** The location of the study area

#### 4.2.2. Field data

The LAI measurements were obtained for each sample plant species in a designed 10 m × 10 m subplot within the 30 m × 30 m plot. The hand-held plant canopy analyser (LAI-2200C) was used to measure LAI of different vegetation classes across the study area. The LAI-2200C calculates LAI and other canopy structure attributes from radiation measurements made with ‘fisheye’ optical sensor (LICOR, 2018; Brusa and Bunker, 2014). A total of 115 ( $N = 115$ ) samples were collected randomly in the study area for all possible vegetation types and croplands. In addition to LAI, information about the dominant vegetation species within the randomly selected plots was collected. Table 4.1 and figure 4.2 give the descriptive statistics and the frequency of LAI in each category respectively.

**Table 4.1:** Descriptive statistics of the measured field LAI in the current study

	<b><i>N</i></b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Std.dev</b>	<b>Variance</b>
LAI	115	0.3	6.33	2.85	1.32	1.75



**Figure 4.2.** The histogram showing the frequency of LAI distribution per category. In this current study, the highest number of field observations had LAI values of between 2 and 4  $m^2/m^2$ .

#### 4.3.3. Image data

Sentinel-2 imagery acquired on between August and September 2017 were used in this chapter. The imagery were obtained from <https://scihub.copernicus.eu/>, comprising of the thirteen (13) total number of spectral bands, at 10 m (visible-near infrared), 20 m (red edge bands and shortwave infrared bands) and 60 m (coastal, cirrus and water vapour bands). For the purpose of this research, only ten (10) multispectral bands in the visible-near infrared (490 – 842 nm), red edge region (705 – 865 nm) and shortwave infrared region (1610 – 2190 nm) were used to retrieve LAI in the study. These spectral are known to be highly correlated with vegetation biophysical and chemical properties of vegetation such as LAI (Clevers *et al.*, 2017; Masemola *et al.*, 2016), chlorophyll (Clevers and Gitelson, 2013; Li *et al.*, 2018), and biomass (Sibanda *et al.*, 2015; Castillo *et al.*, 2017) mainly in homogenous environments. The Sentinel-2 (A/B) missions offer improved temporal resolution of 5 days, which is very important for high resolution mapping of vegetation phenology and biomass.

The 10 multispectral bands of Sentinel-2 were pre-processed in Sentinel Application Platform (SNAP) toolbox of the European Space Agency (ESA). The atmospheric correction was carried out using Sen2Cor algorithm in SNAP, which uses a large database of look-up tables (LUT) derived using an atmospheric radiative transfer model based on libRadtran1 (Müller-Wilm, 2016). All of the atmospherically corrected spectral bands ( $n = 10$ ) were resampled to 10 m spatial resolution, and stacked together to form individual 10-band imagery. Image pre-

processing was concluded when the individual Sentinel-2 images (scenes) were mosaicked and clipped to match the extent of the study area in Quantum GIS software (QGIS Development Team, 2018).

### 5.2.1. Statistical analysis

The statistical relationship between LAI or effective LAI (LAI<sub>effective</sub>) and remote sensing data was determined by assessing its correlation with vegetation indices (BBVI and NBVI) derived from Sentinel-2 data. Five vegetation indices were tested in this chapter based on their relationship with LAI (Zheng and Moskal, 2009) and other vegetation parameters including foliar nitrogen concentration, which is correlated to green vegetation LAI (Klodd *et al.*, 2016; Delegido *et al.*, 2011;). Both the MTVI<sub>2</sub> and MCARI<sub>2</sub> have shown higher correlation to LAI than the common NDVI and other NIR-centric indices (Haboudane *et al.*, 2004). On the other hand, the WDRVI has shown robust linearity with LAI than the conventional red-NIR indices such as NDVI due to the weighting parameter at NIR spectral region (Gitelson, 2004). In comparison, the two red-edge centric vegetation indices were computed as narrowband indices because of their ability to account for signal saturation at vegetation with LAI > 4 and 5 m<sup>2</sup>.m<sup>-2</sup> (Baret *et al.*, 2007; Mutanga and Skidmore, 2004). A limited number of VIs was selected in order to avoid model over-fit. The list of vegetation indices is shown in table 4.2.

**Table 4.2:** The vegetation indices used in the study, utilizing green, red NIR and red edge spectral regions

Index	Formula	Reference
Modified triangular vegetation index (MTVI <sub>2</sub> )†	$MTVI_2 = \frac{[1.5(R_{842} - R_{560}) - 2.5(R_{660} - R_{560})]}{[(2R_{842} + 1)^2 - (R_{842} - 5R_{660}^{0.5}) - 0.5]^{0.5}}$	Haboudane <i>et al.</i> (2004)
Modified chlorophyll absorption ration index (MCARI <sub>2</sub> )†	$MCARI_2 = \frac{1.2[2.5(R_{842} - R_{660}) - 1.3(R_{842} - R_{560})]}{\sqrt{(2R_{842} + 1)^2 - (6R_{842} - 5\sqrt{R_{660}}) - 0.5}}$	Haboudane <i>et al.</i> (2004)
Wide dynamic range vegetation index (WDRVI)†	$WDRVI = \frac{[\alpha(R_{842}) - R_{660}]}{[\alpha(R_{842}) + R_{660}]}$	Gitelson (2004)




---

<p>Normalized difference red edge1 vegetation index (NDVI<sub>re1</sub>)#</p>	$\text{NDVI}_{\text{re1}} = \frac{R_{842} - R_{705}}{R_{842} + R_{705}}$	<p>Gitelson and Merzlyak (1994)  Gitelson <i>et al.</i> (2003)</p>
<p>Red edge chlorophyll index (CI<sub>red edge</sub>)#</p>	$\text{CI}_{\text{red edge}} = \frac{R_{780}}{R_{705}} - 1$	

---

† denotes BBVI, # denotes NBVI in the red edge region, the value of  $\alpha=0.1$

---

The multiple linear regression model was implemented to correlate derived vegetation indices with field LAI measurements, and to predict the LAI from Sentinel-2 imagery. From the standard field dataset ( $N = 115$ ), approximately 60% ( $n = 69$ ) was used for calibrating and training the model, while the remaining 40% ( $n = 46$ ) of the independent dataset was used for model cross-validation in order to test the robustness and reliability of the training model. The training model involves the stepwise selection of significant vegetation indices which are highly correlated to the measured LAI. The stepwise strategy is crucial for LAI estimation because it minimizes the Akaike's Information Criterion (AIC) value (Hu *et al.*, 2018) while retaining the significant candidate variables for model prediction. Additionally, it handles the multi-collinearity problem that is common with remote sensing variables. The 'MASS' package (Modern Applied Statistics with S) was used in R software (R Development Core Team, 2018) to perform model building and variable selection procedures. The co-efficient of determination ( $R^2$ ) and the root-mean square error (RMSE) values were computed for all models as a measure of accuracy assessment. The analysis of variance (ANOVA) will be conducted between the BBVI and NBVI models, and the combined BBVI-NBVI model to assess predictive model variations and the model significance using a chi-square ( $\chi^2$ ) statistic.

### 4.3. Results

#### 4.3.1. Broadband and narrow-band models

The results in table 4.3 of the initial model indicate the interaction of the VIs with measured LAI. From the table, it can be seen that no input variable is significant ( $p < 0.05$ ), although each of the selected variables have very high correlation with LAI ( $R^2 > 0.79$ ). The stepwise regression model was performed for the BBVI and resulted in the MTVI<sub>2</sub> and MCARI<sub>2</sub> as the significant remote sensing variables, as in equation (4.1):

$$LAI = 30.47 \times MTVI_2 - 19.74 \times MCARI_2 - 1.91 \quad (4.1)$$

In this model, both  $MTVI_2$ ,  $MCARI_2$  variables and  $y$ -intercept were significant at  $p < 0.001$ , with the multiple  $R^2 = 0.76$ . On the other hand, the NBVI model yielded the relationship depicted in equation (4.2):

$$LAI = -0.68 \times CI_{red\ edge} + 12.20 \times NDVI_{re1} + 0.31 \quad (4.2)$$

The narrow-band model was significant at  $p < 0.00001$ , with the multiple  $R^2$  of 0.72. In this model, the model significant variable was  $NDVI_{re1}$  at  $p < 0.000001$ .

**Table 4.3:** The results of the initial combined BBVI and NBVI regression model and the variable correlations to measured LAI

Index	Group	Coefficient	Std. Error	<i>t</i>	<i>p</i>	$R^2$
<i>Intercept</i>	-	0.144	4.55	0.032	0.98	-
$MTVI_2$	BBVI	19.215	14.69	1.311	0.19	0.84
$MCARI_2$	BBVI	-12.897	10.96	-1.177	0.24	0.82
WDRVI	BBVI	1.589	4.711	0.337	0.74	0.83
$NDVI_{re1}$	NBVI	5.521	3.939	1.402	0.16	0.85
$CI_{red\ edge}$	NBVI	-0.986	1.332	-0.740	0.46	0.79

However, comparison between final model and NBVI model has resulted in a  $\chi^2$  test with  $p$  value of 0.038, which is statistically significant. On the other hand, the comparison between the final model and the BBVI model yielded a  $\chi^2$  test with value of  $p > 0.373$  and was not statistically significant.

The final model selected through the stepwise strategy is given by equation (4.1). In other words, the selected BBVI model was also the final model to be used for predictive mapping. In this model, the  $MTVI_2$  was positively correlated to the LAI distribution ( $p < 0.004$ ) while the  $MCARI_2$  was negatively correlated to the LAI distribution ( $p < 0.0001$ ) at the study area. Table 5.4 shows the estimates of the final model employed in this chapter and the significance of each input variable.

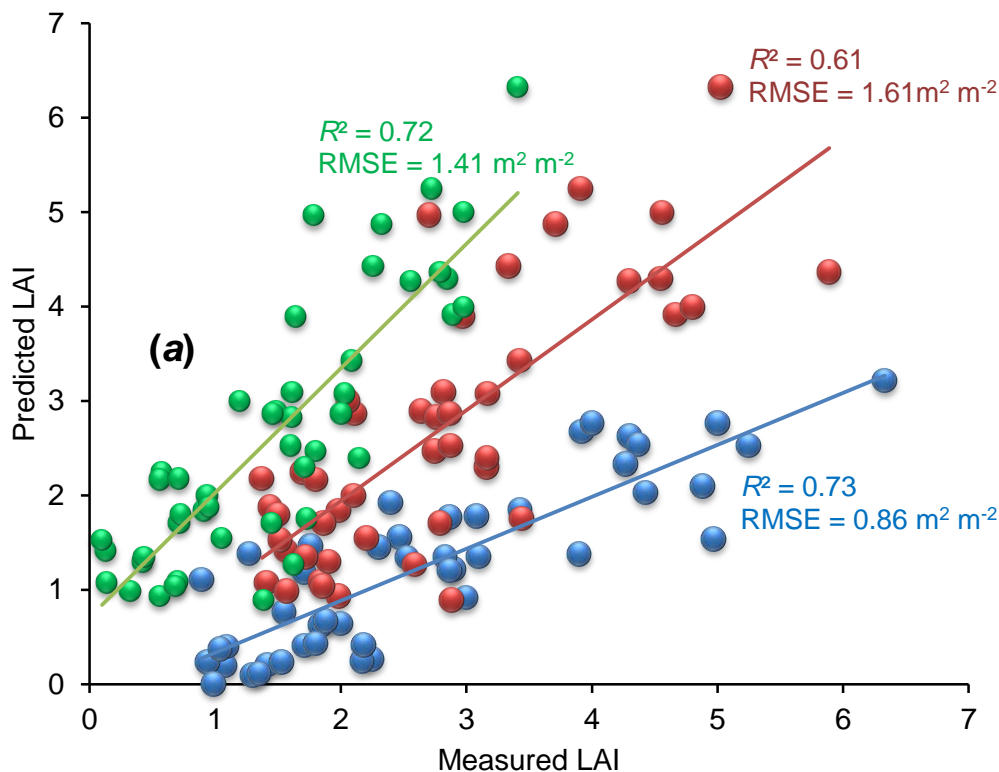
**Table 4.4:** The results of the significant BBVI model used for mapping LAI ( $n = 66$ )

Variable	Coefficient	Std. Error	<i>t</i>	<i>p</i>	$R^2$
<i>Intercept</i>	-1.910	0.510	-3.743	0.0004	-
$MTVI_2$	30.473	7.281	4.185	0.0001	0.84

MCARI <sub>2</sub>	-19.737	6.552	-3.012	0.004	0.82
--------------------	---------	-------	--------	-------	------

### 5.3.1. Accuracy assessment

The independent dataset ( $n=49$ ) was used to cross-validate the predictive accuracy of the regression models. Figure 4.3 shows the correlations between measured and predicted LAI as shown by (a) BSVI model only, (b) final model, and (c) NBVI model. The highest correlation and lowest RMSE were obtained from the final model ( $R^2 = 0.73$ ;  $RMSE = 0.86 \text{ m}^2 \text{ m}^{-2}$ ), while the model with narrow band indices yielded the lowest  $R^2$  (0.61) with high RMSE ( $1.61 \text{ m}^2 \text{ m}^{-2}$ ). The final model was then used for mapping the spatial distribution of LAI using Sentinel-2 dataset. Figure 4.4 shows the spatial distribution of the retrieved LAI across the Vhembe District Municipality.



**Figure 4.3:** Predicted vs. measured LAI using three linear models (a) BSVI model, (b) NBVI model and (c) final model.

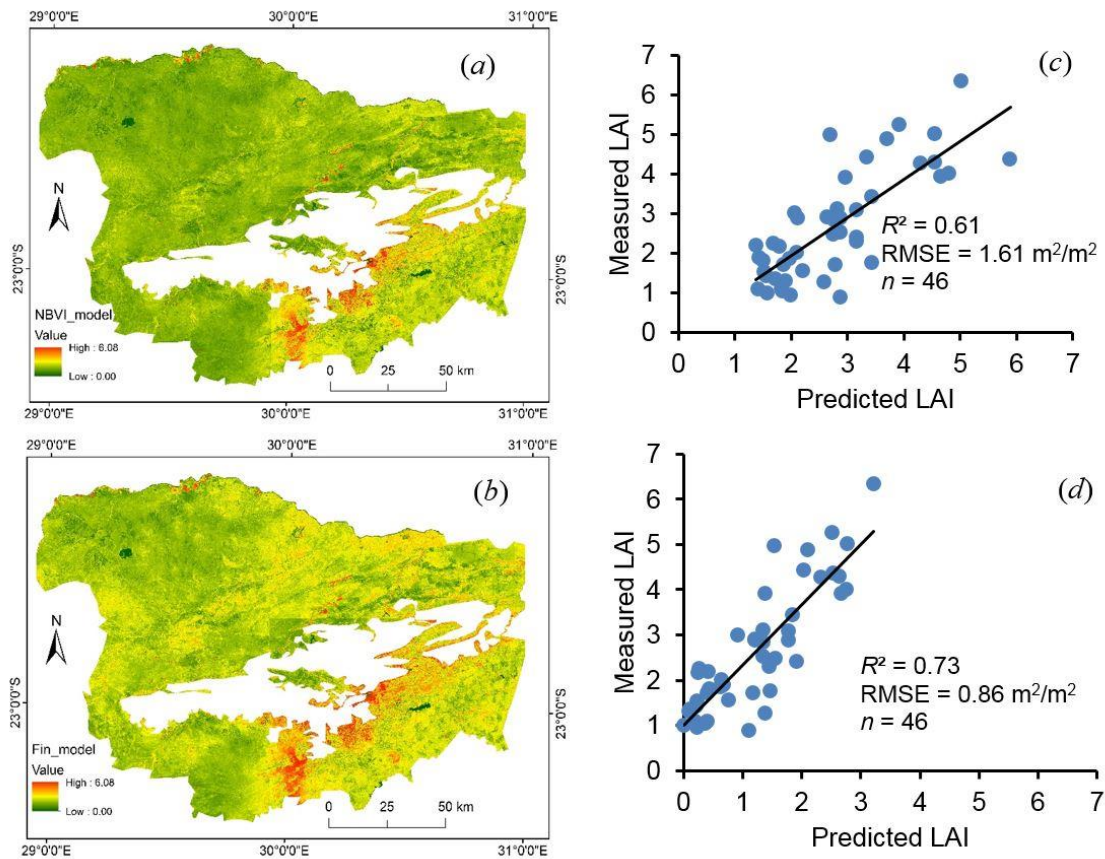
### 4.4. Discussion

The aim of the study was to compare the performance of the broadband vegetation indices with the narrow-band vegetation indices derived from Sentinel-2 instrument to retrieve LAI of



the heterogeneous landscape for malaria predictability. It is in the objective of the study to parameterize field LAI measured from different vegetation types including the compound (e.g. *Dichrostachys cinerea*) and simple leaf vegetation (e.g. *Musa acuminata*). The results of the study showed that both the broadband vegetation indices (BBVI) and the narrow-band (NBVI) have the ability to retrieve the LAI from heterogeneous landscape ( $p < 0.05$ ). However, the current study has showed that the regression model computed from BBVI is more robust than the one computed from NBVI for LAI retrieval. These findings are in agreements with findings by [Clevers et al. \(2017\)](#) who concluded that the broadband vegetation indices at 10m spatial resolution can retrieve LAI with higher accuracy than the narrowband indices, which are usually available at 20 m spatial resolution. The use of either BBVI or the NBVI has shown to be dependent on the distribution of LAI across the heterogeneous landscape. Findings in this chapter show that the BBVI which included MCARI<sub>2</sub>, MTVI<sub>2</sub> and WDRVI yielded higher classification accuracies than the NBVI mainly in the LAI categories of 0-2, 2-4 but underestimated the higher LAI (>5). On the other hand, the NBVI model has shown high sensitivity to LAI greater than 4 (Figures 4.3 and 4.4). The prediction of LAI by final selected stepwise regression model (MCARI<sub>2</sub> and MTVI<sub>2</sub>) has shown that these broadband indices are less affected by the soil background, but suffer signal saturation at moderately higher LAI. This is because MTVI<sub>2</sub> and MTVI<sub>2</sub> have a soil adjustment factor that is otherwise not available in the narrow-band indices used for this chapter ([Jin et al., 2015](#)). Some studies have found that MCARI<sub>2</sub> tends to yield higher prediction of vegetation biophysical characteristics such as the LAI due to the combination of reflectance in the NIR, red and blue wavelengths than other single or two-band indices ([Stagakis et al., 2010](#); [Gu et al., 2015](#)) In contrast, the narrow band indices are have shown ability to account for signal saturation that is common at moderately higher LAI values ([Mutanga and Skidmore, 2004](#)). The performance of BBVI is an encouraging factor especially for retrieval of LAI with satellite datasets that have limited spectral resolution such as IKONOS, CBERS-3/4, and SPOT. It further vindicates the use of Sentinel-2 satellite instruments (S2-A and B) for very high spatial resolution LAI retrieval for multiple applications.





**Figure 4.4:** The results of LAI retrieval from narrow-band vegetation indices (a) vs. the final broadband vegetation indices (b) at the study area. The scatterplots show the respective correlation between measured and predicted LAI for the NBVI (c) and the final BBVI model (d) respectively.

Previous studies have considered validation of the contribution of red edge bands for estimation of LAI in different areas (Frampton *et al.*, 2013; Verrelst *et al.*, 2013). A study by Hermann *et al.* (2011) concluded that, from the simulated Sentinel-2 dataset, the red edge band exhibited higher sensitivity to LAI variations in homogeneous environment than the broad VNIR spectral bands, and has since suggested further testing with real Sentinel-2 data. In the current study, it was found that the contribution of the red edge indices showed rather inferior performance for retrieving LAI when the calibration of such vegetation biophysical parameter is calibrated from vegetation communities of various canopy architectural designs. These findings may have been attributed to aggregation of LAI from different plant species with contrasting canopy structure (foreground), background reflectance (e.g. soil) and phenological conditions in sampling plots.

#### 5.4.1 Spatial mapping and considerations

Overall, the pattern of predicted LAI from Sentinel-2 increases from west (lower LAI values –  $0.0 \text{ m}^2 \text{ m}^{-2}$ ) to east (higher LAI values –  $6.08 \text{ m}^2 \text{ m}^{-2}$ ), and increases also from north to south of the study area (Figure 4.4). On the whole, the eastern half of the study area exhibit high levels of LAI than the western part. This is because the western part of the study area is largely comprised of sparse shrubland mostly with bush-encroaching species such as *D. cinerea*, *Acacia tortilis* and *A. mellifera*, which are characteristic of savannah biome (Munyati *et al.*, 2013). These bush-encroaching species are found in warm semi-arid climatic zones, with different leaf architecture and leaf area, than vegetation in the humid sub-tropical climate within the African savannah (Roques *et al.*, 2001; Oldeland *et al.*, 2010). Although the LAI was successfully mapped in this chapter, it is albeit crucial to consider the limitations that arise when retrieving LAI from Sentinel-2 dataset: Firstly, the regression model could not extrapolate LAI values beyond the range of the training dataset, thereby not sufficiently giving indication of ‘true’ LAI distribution especially in irrigated agricultural croplands. Secondly, due to limited number of samples in collected for LAI greater than 5, the LAI distribution maps may be biased towards lower LAI values/categories hence the underestimation of high LAI resulting from the MCARI<sub>2</sub>-LAI and MTVI<sub>2</sub>-LAI model. Finally, it should be noted that the retrieval of LAI in the current chapter was done for imagery obtained in a single time period, which considerably affects the results of the study. It is thus recommended to test the performance of indices across multi-seasonal imagery to capture the dynamics of vegetation phenology and other biophysical characteristics impacting vegetation leaf area.

#### 4.5. Conclusions

In summary, this chapter has demonstrated the ability of Sentinel-2 dataset for retrieving LAI calibrated from heterogeneous vegetation. The study has shown that the broad-band vegetation indices of Sentinel-2 yielded higher mapping accuracy ( $R^2 = 0.73$ ; RMSE =  $0.86 \text{ m}^2 \text{ m}^{-2}$ ;  $p < 0.05$ ) than the narrow-band vegetation indices ( $R^2 = 0.61$ ; RMSE =  $1.61 \text{ m}^2 \text{ m}^{-2}$ ;  $p < 0.05$ ), with the statistically significant difference between the two models ( $p < 0.038$ ). The model derived from three-band MCARI<sub>2</sub> and MTVI<sub>2</sub> with soil-adjustment factor was eventually used for spatial modelling of the LAI in the Vhembe District, demonstrating that superior LAI estimation can already be achieved at 10 m spatial resolution (Verrelst *et al.*, 2013). It is therefore recommended that these results be interpreted with caution because of the number of sample and the structure of the data used to derive the LAI. These results can further be validated by considering the multi-seasonal approach to LAI mapping which was not

considered in the current study. The results from this study indicate that the broadband vegetation indices are better predictors of the leaf area, which is one of the most important components involved in the biology of vector pathogen transmissivity. Remote sensing techniques thus enable us to effectively quantify this crucial vegetation biophysical characteristic, which is an important factor in the study of malaria distribution. The distribution maps derived from remote sensing show that the areas with low LAI derived from S-2 BBVI correspond to areas of low malaria risk, while the opposite is the case with the areas of high predicted LAI values.

Communities in the rural areas of the Vhembe District Municipality depend on various ethnobotanical plants as crucial sources for primary healthcare against malaria (Phaswane and Masevhe, 2018). One of the most commonly used species is the *Lippia javanica* (lemon bush) – a shrub species belonging to the Verbenaceae family and grows up to 2 m in height (Xaba and McVay, 2010). The *L. javanica* is used, among other functions, as an alternative for western medicine that repels mosquitoes, thereby controlling malaria infections in the areas. The following chapter will be presenting remote sensing methods for mapping the distribution of *L. javanica* at the study area.

In summary, this chapter presented the novel approach to retrieving the LAI of vegetation within the study area. It was found that the LAI derived from the common visible-near infrared bands of Sentinel-2 is more accurate than the one derived at narrow-band. This chapter addresses objective 2 of the study.

#### 4.6. References

- Afrane Y.A., Klinkenberg E., Drechsel P., Owusu-Daaku K., Garms R., Kruppa T. (2004). Does irrigated urban agriculture influence the transmission of malaria in the city of Kumasi, Ghana? *Acta Tropica*, 89:125–34.
- Asano T., Kosugi, Y., Uto K., Kosaka N., Odagawa S., Oda K. (2009). Leaf area index estimation from hyperspectral data using group division method. *IEEE International Geoscience and Remote Sensing Symposium*, Cape Town, 2009, pp. IV-817-IV-820. doi: 10.1109/IGARSS.2009.5417502.

- Asner P.G., Martin R.E., Anderson C.B., Knapp D.E. (2015). Quantifying forest canopy traits: Imaging spectroscopy versus field survey. *Remote Sensing of Environment*, 158: 15 – 27.
- Baret F., Hagolle O., Geiger B., Bicheron P., Miras B., Huc M., Berthelot B., Nino F., Weiss M., Samain O., Roujean J.L., Leroy M. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. *Remote Sensing of Environment*, 110(3): 275 – 286.
- Borzuchowski J., Schultz K. (2010). Retrieval of Leaf Area Index (LAI) and Soil Water Content (WC) Using Hyperspectral Remote Sensing under Controlled Glass House Conditions for Spring Barley and Sugar Beet. *Remote Sensing 2*: 1702 – 1721.
- Breda N.J.J.(2003). Ground-based measurements of leaf area index: A review of methods, instruments and current controversies. *Journal of Experimental Botany*, 5: 2403–2417.
- Breunig F.M., Galvão L.S., Formaggio A.R., Epiphanyo J.C.N. (2011). Directional effects on NDVI and LAI retrievals from MODIS: A case study in Brazil with soybean. *International Journal of Applied Earth Observation and Geoinformation*, 13(1): 34 – 42.
- Brusa A., Bunker D. (2014). Increasing the precision of canopy closure estimates from hemispherical photography: Blue channel analysis and under-exposure. *Agricultural and Forest Meteorology*, 195-196:102 – 107.
- Carmona F., Rivas R., Fonnegra D.C. (2015). Vegetation Index to estimate chlorophyll content from multispectral remote sensing data. *European Journal of Remote Sensing*, 48: 319 – 326.
- Castillo J.A.A., Apan A.A., Maraseni T.N., Salmo III S.G. (2017). Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134: 70 – 85.
- Clark M.L., Roberts D.A., Clark D.B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96: 375 – 398.
- Clark P.J., Lawes M.J., Murphy B.P., Russell-Smith J., Nano C.E.M., Bradstock R., Enright N.J., Fontaine J.B., Gosper C.R., Radford I., Midgley J.J., Gunton R.M. (2015). A synthesis of postfire recovery traits of woody plants in Australian ecosystems. *Science of the Total Environment*, 534: 31 – 42.



- Clevers J.G.P.W., Gitelson A.A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation*, 23: 344 – 351.
- Clevers J.G.P.W., Kooistra L., van den Brande M.M.M. (2017). Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing*, 9(5): 405. <https://doi.org/10.3390/rs9050405>.
- Cole B., McMorro J., Evans M. (2014). Spectral monitoring of moorland plant phenology to identify a temporal window for hyperspectral remote sensing of peatland. *ISPRS Journal of Photogrammetry & Remote Sensing*, 90: 49 – 58.
- Darvishzadeh R., Atzberger C., Skidmore A.K. Hyperspectral vegetation indices for estimation of leaf area index. In: *Proceedings of ISPRS Commission VII Mid-term Symposium*, Enschede, the Netherlands, 8–11 May 2006
- Darvishzadeh R., Skidmore A., Schlerf M., Atzberger C. (2008b). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, 112:2592 – 2604.
- Darvishzadeh R., Skidmore A., Schlerf M., Atzberger C., Corsia F., Cho M. (2008a). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry & Remote Sensing*, 63:409–426.
- Delegido J., Verrelst J., Alonso L., Moreno J. (2011). Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors*, 11: 7063 – 7081.
- Delegido J., Verrelst J., Meza C.M., Rivera J.P., Alonso L., Moreno J. (2013). A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *European Journal of Agronomy* 46: 42 – 52.
- Dewald J.R., Fuller D.O., Muller G.C., Beier J.C. (2016). A novel method for mapping village-scale outdoor resting microhabitats of the primary African malaria vector, *Anopheles gambiae*. *Malaria Journal*, 15:489. DOI 10.1186/s12936-016-1534-9.
- Din M., Zheng W., Rashid M., Wang S., Shi Z. (2017). Evaluating Hyperspectral Vegetation Indices for Leaf Area Index Estimation of *Oryza sativa* L. at Diverse Phenological Stages. *Frontiers in Plant Science*, 8(820): doi: 10.3389/fpls.2017.00820.
- Dorigo W., Richter R., Baret F., Bamler R., Wagner W. (2009). Enhanced Automated Canopy Characterization from Hyperspectral Data by a Novel Two Step Radiative Transfer Model Inversion Approach. *Remote Sensing*, 1: 1139 – 1170.





- Frampton W., Dash J., Watmough G., Milton E.J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013: 83 – 92.
- Gausman H.W. (1982). Visible light reflectance, transmittance, and absorptance of differently pigmented cotton leaves. *Remote Sensing of Environment*, 13: 233 – 238.
- Gholizadeh A., Kopaekova V., Rogass C., Mielke C., Mišurec J. Extracting forest canopy characteristics from remote sensing imagery: Implications for Sentinel-2 mission. *Proceedings of Living Planet Symposium (2016)*, Prague, Czech Republic, 9 – 13 May 2016 (ESA SP-740, August 2016).
- Gitelson A.A. (2004). Wide Dynamic Range Vegetation Index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology*, 161:165–173. doi:10.1078/0176-1617-01176
- Gitelson A.A., Gritz Y., Merzlyak M.N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160: 271–282
- Gitelson A.A., Keydan G.P., Merzlyak M.N. (2006). Three-Band Model for Non-invasive Estimation of Chlorophyll Carotenoids and Anthocyanin Contents in Higher Plant Leaves. *GeoPhysical Research Letters*, 33: L11402, doi:10.1029/2006GL026457.
- Gitelson A.A., Merzlyak M.N. (1994). Spectral reflectance changes associated 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.A., Merzlyak M.N. (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research*, 22(5): 689 – 692.
- Gitelson A.A., Wardlow B.D., Keydan G.P., Leavitt B. (2007). An evaluation of MODIS 250-m data for green LAI estimation in crops. *Geophysical Research Letters*, 34(L20403): 1 – 4.
- Glenn E.P., Huete A.R., Nagler P.L., Nelson P.G. (2008). Relationship Between Remotely-sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape. *Sensors*, 8(4): 2136–2160.
- Gu W.D., Novak R.J. (2005). Habitat-based modeling of impacts of mosquito larval interventions on entomological inoculation rates, incidence, and prevalence of malaria. *American Journal of Tropical Medicine Hygiene* 73:546–52.

- Gu Z.J., Sanchez-Azofeifa G., Feng J., Cao S. (2015). Predictability of leaf area index using vegetation indices from multiangular CHRIS/PROBA data over eastern China. *Journal of Applied Remote Sensing*, 9(1): 096085. <https://doi.org/10.1117/1.JRS.9.096085>
- Haboudane D., Miller J.R., Temblay N., Zarco-Tejada P.J., Dextraze L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, 81(2/3): 416 – 426.
- Haboundane D., Miller j.R., Pattey E., Zarco-Tejada P.J., Strachan I.B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3): 337 – 352.
- Hansen P.M., Schjoerring K.J. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86: 542 – 553.
- Hermann I., Pimstein A., Karnieli A., Cohen Y., Alchanatis V., Bonfil D.J. (2011). LAI assessment of wheat and potato by VEN $\mu$ S and Sentinel-2 bands. *Remote Sensing of Environment*, 115: 2141-2151.
- Houborg R., Boegh E. (2008). Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sensing of Environment*, 112(1):186 – 202.
- Houborg R., Fisher J.B., Skidmore A.K. (2015). Advances in remote sensing of vegetation function and traits. *International Journal of Applied Earth Observation and Geoinformation*, 43: 1–6.
- Hu Q., Ma Y., Xu B., Song Q., Tang H., Wu W. (2018). Estimating Sub-Pixel Soybean Fraction from Time-Series MODIS Data Using an Optimized Geographically Weighted Regression Model. *Remote Sensing*,
- Hunt Jr. E.R, Doraiswamy P.C., McMurtrey J.E., Daughtry C.S.T., Perry E.M., Akhmedov B. (2013). A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International Journal of Applied Earth Observation and Geoinformation*, 21:103 – 112.
- Jin X., Yang G., Xu X., Yang H., Feng H., Li Z., Shen J., Zhao C., Lan Y. (2015). Combined Multi-Temporal Optical and Radar Parameters for Estimating LAI and



Biomass in Winter Wheat Using HJ and RADARSAR-2 Data. *Remote Sensing*, 7(10): 13251 – 13272.

- Jordan C. F. (1969). Derivation of leaf area index from quality of light on the forest floor. *Ecology*, 50, 663 – 666.
- Killeen G.F., Govella N.J., Lwetoijera D.W., Okumu F.O. (2016). Most outdoor malaria transmission by behaviourally-resistant *Anopheles arabiensis* is mediated by mosquitoes that have previously been inside houses. *Malaria Journal*, 15:255. DOI: 10.1186/s12936-016-1280-z.
- Kira O., Linker R., Gitelson A. (2015). Non-destructive estimation of foliar chlorophyll and carotenoid contents: Focus on informative spectral bands. *International Journal of Applied Earth Observation and Geoinformation*, 38: 251 – 260.
- Klodd A.E., Nippert J.B., Ratajczak Z., Waring H., Phoenix G.K. (2016). Tight coupling of leaf area index to canopy nitrogen and phosphorus across heterogeneous tallgrass prairie communities. *Oecologia*, 182: 889 – 898.
- Li C., Zhu X., Wei Y., Cao S., Guo X., Yu X., Chang C. (2018). Estimating apple tree canopy chlorophyll content based on Sentinel-2A remote sensing imaging. *Scientific Reports*, 8: 3756.
- Li F., Miao Y., Feng G., Yuan F., Yue S., Gao X., Liu Y., Liu B., Ustin S.L., Chen. (2014). Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. *Field Crops Research*, 157: 11 – 123.
- Lichtenthaler H.K., Gitelson A., Lang M. (1996). Non-destructive determination of chlorophyll content of leaves of a green and an aurea mutant of tobacco by reflectance measurements. *Journal of Plant Physiology*, 148: 483–493
- Limpopo State of the Environment Report-Phaswe (LSER)(2004). Accessed on: 23-02-2018.  
[http://soer.deat.gov.za/dm\\_documents/Limpopo\\_State\\_of\\_environment\\_reporting\\_phase\\_1\\_UrZKq.pdf](http://soer.deat.gov.za/dm_documents/Limpopo_State_of_environment_reporting_phase_1_UrZKq.pdf)
- Liu K., Zhou Q., Wu W., Xia T., Tang H. (2016). Estimating the crop leaf area index using hyperspectral remote sensing: a review. *Journal of Integrative Agriculture*, 15(2): 475–491.
- Main R., Cho M.A., Mathieu R., O Kennedy M.M., Ramoelo A., Koch S. (2011). An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS Journal of Photogrammetry and Remote Sensing* 66: 751–761.

- Marr I.L., Suryana N., Lukulay P., Marr M.I. (1995). Determination of chlorophyll a and b by simultaneous multi-component spectrophotometry. *Fresenius Journal of Analytical Chemistry*, 352: 456 – 460.
- Masemola C., Cho M.A., Ramoelo A. (2016). Comparison of Landsat 8 OLI and Landsat 7 ETM+ for estimating grassland LAI using model inversion and spectral indices: case study of Mpumalanga, South Africa. *International Journal of Remote Sensing*, 37(18): 4401 – 4419..
- Mayagaya V.S., Nkwengulila G., Lyimo I.N., Kihonda J., Mtambala H., Ngonyani T., Russell T.L., Furguson H.M. (2015). The impact of livestock on the abundance, resting behaviour and sporozoite rate of malaria vectors in southern Tanzania. *Malaria Journal*, 14(17): <https://doi.org/10.1186/s12936-014-0536-8>.
- Milton E.J., Schaepman M.E., Anderson K., Kneubühler M., Fox N. (2009). Progress in field spectroscopy. *Remote Sensing of Environment*, 113:92 – 109.
- Moro M.F., Silva I.A., de Araujo F.S., Lughadha E.N., Meagher T.R., Martins F.R. (2015). The Role of Edaphic Environment and Climate in Structuring Phylogenetic Pattern in Seasonally Dry Tropical Plant Communities. *PLoS One*, 10(3):e01119166.
- Müller-Wilm U. (2016). Sentinel-2 MSI – Level-2A Prototype Processor Installation and User Manual. S2PAD-VEGA-SUM-0001, Issue (2.2).
- Munyati C., Economon E.B., Malahlela O.E. (2013). Effect of canopy cover and canopy background variables on spectral profiles of savanna rangeland bush encroachment species based on selected Acacia species and *Dichrostachys cinerea* at Mokopane, South Africa. *Journal of Arid Environments*, 94 (2013): pp. 121–126.
- Mutanga O., Adam E., Adjorlolo C., Abdel-Rahman E.M. (2015). Evaluating the robustness of models developed from field spectral data in predicting African grass foliar nitrogen concentration using WorldView-2 image as an independent test dataset. *International Journal of Applied Earth Observation and Geoinformation*, 34: 178 – 187.
- Mutanga O., Skidmore A.K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19): 3999 – 4014
- Oldeland J., Dorigo W., Wesuls D., Jurgens N. (2010). Mapping Bush Encroaching Species by Seasonal Differences in Hyperspectral Imagery. *Remote Sensing*, 2: 1416 – 1438.
- Opti-Sciences Incorporated (2017). CCM-300 Chlorophyll Content Meter - for very small leaves and difficult samples. 8 Winn Avenue, Hudson, NH03051, USA.

- Paaijmans K.P., Thomas M.B. (2011). The influence of mosquito resting behaviour and associated microclimate for malaria risk. *Malaria Journal*, 10:183. <https://doi.org/10.1186/1475-2875-10-183>.
- Peng D., Jiang Z., Huete A.R., Ponce-Campos G.E., Nguyen U., Luvall J.C. (2013). Response of Spectral Reflectances and Vegetation Indices on Varying Juniper Cone Densities. *Remote Sensing*, 5: 5330 – 5345.
- Phaswane M.C., Masevhe N.A. (2018). Investigation of medicinal plants for management of malaria at Tshakhuma community, Makhado Municipality, Vhembe District, Limpopo province, Republic of South Africa. *South African Journal of Botany*, 115: 327.
- QGIS Development Team (2018). QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org> .
- R Development Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- R Development Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Reason C.J.C., Keibel A. (2004). Tropical cyclone *Eline* and its unusual penetration and impacts over the southern African mainland. *Weather Forecast*, 19(5):789–805.
- Ricotta E.E., Frese S.A., Choobwe C., Louis T.A., Shiff C.J. (2014). Evaluating local vegetation cover as a risk factor for malaria transmission: a new analytical approach using ImageJ. *Malaria Journal*, 13(94): DOI: 10.1186/1475-2875-13-94.
- Rondeaux G., Steven M., Baret F. (1994). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55: 95 – 107.
- Rouques K.G., O'Connor T.G., Watkinson A.R. (2001). Dynamics of shrub encroachment in an African savanna: relative influences of fire, herbivory, rainfall and density dependence. *Journal of Applied Ecology*, 38: 268 – 280.
- Rouse J.W., Haas R.H., Schell J.A., Deering D.W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. In: S.C. Freden and M.A. Becker, editors, Third Earth Resources Technology Satellite Symposium, Greenbelt, MD. 10–14 Dec. 1973. NASA, Washington, DC. p. 309–317.
- Rutherford M.C., Mucina L., Lötter M.C., Bredenkamp G.J., Smit C., Scott-Shaw R., Jacobus .H.L., Hoare D.B., Goodman P.S., Bezuidenhout H., Scott L., Ellis F., Powrie



- L.W., Siebert F., Mostert T.H., Henning B.J., Venter C.E., Camp K.G.T, Siebert S.J., Matthews W.S., Burrows J.E., Dobson L., van Rooyen N., Schmidt E., Winter P.J.D., Johann du Preez P., Ward R.A., Williamson S., Hurte P.J.H. (2006). Savanna biome. *South African National Biodiversity Institute*, p438-538.
- Ryu C., Suguri M., Umeda M. (2009). Model for predicting the nitrogen content of rice at panicle initiation stage using data from airborne hyperspectral remote sensing. *Biosystems Engineering*, 104 465–475.
  - Sampson P.H., Zarco-Tejada, P.J., Mohammed, G.H., Miller, J.R., Noland, T.L., 2003. Hyperspectral remote sensing of forest condition: estimating chlorophyll content in tolerant hardwoods. *Forest Science*, 49: 381–391.
  - Sibanda M., Mutanga O., Rouget M. (2015). Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 110: 55 – 65.
  - Sims D.A., Gamon J.A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81 (2/3): 337 – 354.
  - Stagakis S., Markos N., Sykioti O., Kyparissis A. (2010). Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: An application on a *Phlomis fruticosa* Mediterranean ecosystem using multiangular CHRIS/PROBA observations. *Remote Sensing of Environment*, 114(5): 977 – 994.
  - Sun L., Gao F., Anderson M.C., Kustas W.P., Alsina M.M., Sanchez L., Sams B., McKee L., Dulaney W., White W.A., Alfieri J.G., Prueger J.H., Melton F., Post K. (2017). Daily Mapping of 30 m LAI and NDVI for Grape Yield Prediction in California Vineyards. *Remote Sensing*, 9(4): 317. doi:10.3390/rs9040317.
  - Tong A., He Y. (2017). Estimating and mapping chlorophyll content for a heterogeneous grassland: Comparing prediction power of a suite of vegetation indices across scales between years. *ISPRS Journal of Photogrammetry and Remote Sensing*, 126: 146 – 167.
  - Verrelst J., Rivera J.P., Moreno J., Camps-Valls G. (2013). Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86: 157 – 167.
  - Viña A., Gitelson A.A., Nguy-Robertson A.L., Peng Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, 115: 3468 – 3478.



- Wu C., Niu Z., Tang Q., Huang W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, 148: 1230–1241.
- Xaba P., McVay R. (2010). Gardening with indigenous traditionally-useful plants. *The Lemon Bush*. South African National Biodiversity Institute (SANBI). Accessed: 29 October 2018.  
<http://www.botanicalsociety.org.za/ProjectsAndActivities/Useful%20Plants/13%20The%20Lemon%20Bush%20Sept.%202010.pdf>
- Xavier A.C., Vettorazzi C.A. (2004). Mapping leaf area index through spectral vegetation indices in a subtropical watershed. *International Journal of Remote Sensing*, 25(9): 1661 – 1672.
- Xie Q., Huang W., Liang D., Chen P., Wu C., Yang G., Zhang J., Huang L., Zhang D. (2014). Leaf Area Index Estimation Using Vegetation Indices Derived From Airborne Hyperspectral Images in Winter Wheat. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8): 3586 – 3594.
- Yang X., Yu Y., Fan W. (2015). Chlorophyll content retrieval from hyperspectral remote sensing imagery. *Environmental Monitoring and Assessment*, 187(7):456. doi: 0.1007/s10661-015-4682-4.
- Yoder B.J., Pettigrew-Crosby R.E. (1995). Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. *Remote Sensing of Environment*, 5(3): 199 – 211.
- Zheng G., Moskal L.M. (2009). Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors*, 9(4): 2719 – 2745.



## CHAPTER 5<sup>4</sup>:

# Evaluating Efficacy of Landsat-Derived Environmental Covariates for Predicting Malaria Distribution in Rural Villages of Vhembe District, South Africa

Oupa E. Malahlela <sup>a,b\*</sup>, Olwoch J.M. <sup>a,b</sup> Adjorlolo C <sup>b</sup>

<sup>a</sup> *Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa*

<sup>b</sup> *South African National Space Agency (SANSA), Earth Observation Directorate, Pretoria 0001, South Africa*

***This paper relates to objectives 3 of the thesis.***

---

<sup>4</sup>This chapter is based on the manuscript titled “*Evaluating Efficacy of Landsat-Derived Environmental Covariates for Predicting Malaria Distribution in Rural Villages of Vhembe District, South Africa*” EcoHealth: <https://doi.org/10.1007/s10393-017-1307-0>

## Abstract

Malaria in South Africa is still a problem despite existing efforts to eradicate the disease. In the Vhembe District Municipality (VDM) malaria prevalence is still high, with mean incidence rate of 328.2 per 100 000 persons/year. This chapter aimed at evaluating environmental covariates, such as vegetation moisture and vegetation greenness, associated with malaria vector distribution for rapid and efficient disease management and control. The 2005 epidemiological data combined with Landsat 5 ETM were used in this chapter. A total of 9 remotely-sensed covariates were derived while pseudo-absences in the ratio of 1:2 (presence/absence) were generated at buffer distances of 0.5-20 km from known presence locations. A stepwise logistic regression model was applied to analyse spatial distribution of malaria. A buffer distance of 10 km yielded the highest classification accuracy of 82% at a threshold of 0.9. This model was significant ( $p < 0.05$ ) and yielded a deviance ( $D^2$ ) of 36%. The significantly positive relationship ( $p < 0.05$ ) between SAVI and malaria distribution at all buffer distances suggests that malaria vector (*Anopheles arabiensis*) prefer productive and greener vegetation. The significant negative relationship between water/moisture index ( $a_1$  index) and malaria distribution in buffer distances of 0.5 km, 10 km and 20 km suggest that malaria distribution increases with a decrease in shortwave reflectance signal. The study has shown that suitable habitats of malaria vectors are generally found within a radius of 10km in semi-arid environments and this insight can be useful to aid efforts aimed at eradicating malaria by 2020 in South Africa. The study has also demonstrated the utility of Landsat data and the ability to extract environmental conditions which favour the distribution of malaria vector (*An. arabiensis*) such as the canopy moisture content in vegetation, which serves as a surrogate for rainfall.

**Keywords:** Vhembe District Municipality; malaria; SAVI; Landsat 5

---

## 5.1. Introduction

The global malaria infection rates are a public health concern, with over 210 million cases reported across the world in 2015. It is reported approximately 90% of all malaria deaths occur in Africa in the year 2015 (WHO, 2016). In Africa, malaria is commonly transmitted through the bite of female *Anopheline* mosquitoes on humans. Malaria cases are due to *Plasmodium falciparum* pathogen, which utilizes humans as a natural intermediate host. In South Africa, malaria is endemic to low-altitude areas of the northern and eastern parts of KwaZulu-Natal, Mpumalanga and Limpopo provinces. The high malaria transmission is invariably seasonal and is often limited to warm and rainy summer months (Craig *et al.*, 1999). In Vhembe District Municipality, Limpopo province, high malaria cases are reported, with over 2-3 incidences per 1000 population at district level (Raman *et al.*, 2016). South Africa is one of the countries that pledged to eradicate malaria by 2020, and efforts are currently made to realize this intervention for zero-malaria country. These efforts include the implementation of various measures to



achieve a primary milestone that is characterized by four phases. These phases are: (i) controlling malaria to less than 5 positive cases per 1000 persons, (ii) pre-elimination stage (less than 1 case per 1000 persons at risk per year), (iii) complete elimination (no transmission) and (iv) prevention of re-introduction of malaria diseases (NDoH, 2010; Maharaj *et al.*, 2012). The ultimate goal is to achieve zero malaria transmission in a year in malaria prone countries such as South Africa. In order to achieve this, it is crucial to collect the relevant data regarding the occurrence of *P. falciparum* species in endemic areas. This will in turn, assist in efforts aimed at detecting, controlling, managing and eradicating malaria.

Mapping the distribution of *P. falciparum* involves the knowledge of potential habitats of *Anopheles* species and records confirming malaria presence/absence in a particular area or region. This exercise often employs field surveys where health-care workers record areas of malaria presence, which can range from a single household to a regional scale. Nonetheless, this method of data collection takes into account the locations and the frequencies of malaria presence and seldom includes environments where malaria is absent. This poses inherent technical challenges when attempting to model the spatial distribution of malaria, since the data collected may be statistically 'incomplete'. However, a number of species distribution models incorporating the presence-data only or presence-absence data exist as explained in the literature (Phillips *et al.*, 2006) and have been used over the years. Moreover, the availability of presence-absence data is important for modelling the distribution of *P. falciparum*, depending of the statistical method applied for mapping.

Several species distribution models (SDM's) have been widely used in ecology and in epidemiological studies alike. The premise of these models is two-fold: those that use presence-only data, and those that require both presence and absence data (Barbet-Massin *et al.*, 2012). Perhaps one of the few presence-only SDM's is rectilinear envelope including BIOCLIM (Busby, 1991). Some models, although categorised as presence-only model such as the maximum (Maxent), require background pseudo-absence data to be fully functional (Phillips *et al.*, 2006). In addition to Maxent, the Genetic Algorithm for Rule-Set Prediction (GARP) also requires the insertion of pseudo-absence, i.e. 0, to depict areas where the species is unlikely to occur (Stockwell and Peters, 1999). On the other hand, most methods for predicting malaria rely on the availability of both presence-absence records of malaria pathogen. The common example of such methods is logistic regression model, either univariate or multivariate, and boosted regression tree (Clennon *et al.*, 2010; Kleinschmidt *et al.*, 2000; Sinka *et al.*, 2010). To ensure accurate mapping of potential malaria distribution in

instances where absence-data is missing, deriving pseudo-absences for use in regression and mapping can be an effective alternative. One of the advantages of employing presence/absence models for predicting malaria is that they make it possible to assess the precision of map and significance of covariates, while allowing for quantification of errors of estimations (Kazembe *et al.*, 2006).

In cases where the absence data is not readily available, generating pseudo-absences for mapping of malaria has been considered as an alternative by many researchers (Sinka *et al.*, 2010; Nmor *et al.*, 2013; Alimi *et al.*, 2015). Most of these studies that utilize pseudo-absences for malaria mapping are focused on regional scales, with minimal emphasis placed on interactions at local scales (Tonnang *et al.*, 2010). It has been established that pseudo-absences generated very close to presence cases may fall on true unidentified presences, while those generated randomly very far from presences may result in a model defining coarser geographical differences rather than fine-scale variables (Chefaoui and Lobo, 2007). The coarse scale analysis is usually as a result of objective to focus on efforts to control malaria epidemic on larger areas, thereby obscuring interactions of local environmental covariates. For example, Sinka *et al.* (2010) studied the distribution of malaria in the Americas at buffer range of 100 km-1500 km from known presence locations. The findings showed rather a general depiction of malaria distribution on a coarser scale, and not necessarily the local malaria distribution pattern. In contrast, Nmor *et al.* (2013) had effectively predicted malaria vector breeding habitats using the pseudo-absences generated at the spatial distance of greater than 50 m from the presence locations. Generating the pseudo-absences for a particular study of interest partly depends on the scale of available data and the overall study objective. A large body of literature exists, in which the generation of pseudo-absences was performed along environmental and topographic gradients such as rainfall, roads, slopes and distance from rivers (Ahmed, 2014; Machault *et al.*, 2011; Zhou *et al.*, 2012). However, can the derivation of pseudo-absences along satellite data gradient shed light in describing the extent of malaria distribution? If so, how far should pseudo-absences be generated from known presence locations, especially in semi-arid rural villages which are located very close to each other (<20 km radius)?

Satellite data have been extensively used for predicting malaria vector distribution globally (Adeola *et al.*, 2016; Tonnang *et al.*, 2010; Alimi *et al.*, 2015; Omumbo *et al.*, 2002). The efficiency of mapping is achievable considering the fact that satellite/remote sensing data are available at various spatial (0.5 m-1000 km), temporal (daily - yearly) and spectral scales

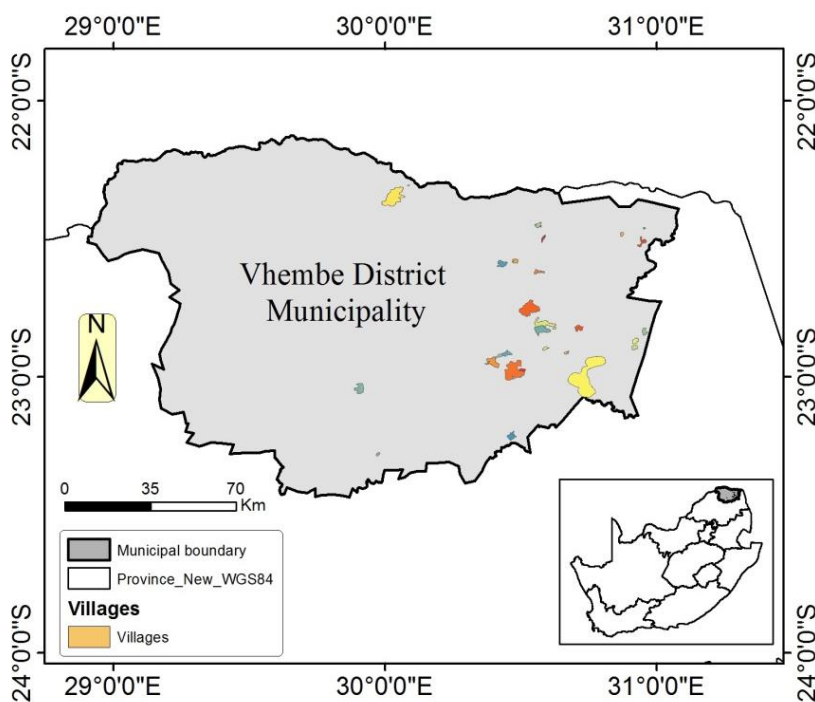
(multispectral - hyperspectral). In Africa, however, a few attempts were made to predict malaria vector breeding sites based on remote sensing, environmental and topographic datasets. For example, [Clennon \*et al.\* \(2010\)](#) employed Landsat 5 and general linear models to characterize mosquito breeding habitats in Zambia. In South Africa the use of satellite data for malaria mapping is extremely limited, contributing to  $\leq 2\%$  of total malaria research ([Adeola \*et al.\*, 2015](#)). Because of this limitation, many local aspects of malaria-environment interactions in South Africa are not well-understood or not adequately quantified. These environmental aspects include vegetation cover and moisture parameters which play a role in malaria transmission. In addition, there is very limited application of satellite technology in mapping malaria especially in malaria-endemic areas of South Africa (Mpumalanga, KwaZulu-Natal and Limpopo provinces). To the best of authors' knowledge, no study was found in literature that sought to assess impact of pseudo-absences generated for medium scale malaria prediction by use of remote sensing covariates. In the Limpopo province of South Africa, malaria prevalence continues to exert pressures of health care services and financial investments associated with disease control and monitoring, particularly in the Vhembe District Municipality. This is because the area experiences high malaria prevalence, with mean incidence rate of approximately 328.2 per 100 000 persons/year ([Gerritsen \*et al.\*, 2008](#)). Therefore, this chapter aimed at evaluating effects of Landsat-derived environmental covariates for predicting malaria distribution in semi-arid rural villages of Vhembe District, South Africa, at buffer distances of 0.5 km to 20 km. This study was the first of its kind, in support of initiatives aimed at eradicating malaria by 2020. Landsat satellites have been in operation since 1972 and provide a desirable temporal, spatial and spectral coverage that are necessary to study the seasonality patterns of a malaria epidemic.

## 5.2. Methods

### 5.2.1. Study area

The rural villages of Vhembe District Municipality in Limpopo Province of South Africa are the ideal candidates to test the study objective. The study area is located at the center geographic coordinates of 23°40' S and 30°00' E (Figure 5.1). It comprises of varying topography, with diverse floral and faunal biodiversity. It receives annual summer rainfall of 820 mm ([Mpandeli, 2014](#)), with Soutpansberg Mountain modifying geographical rainfall patterns ([Kabanda and Munyati, 2010](#)). The north-western part of Vhembe District is characterized by semi-arid conditions, while the south-eastern part experiences subtropical conditions. The Vhembe District Municipality has a population of more than 1.3million people ([StatsSA, 2016](#)), who

predominantly reside in rural villages. Data from 28 villages were used for the study. The area covering these villages has recorded mean malaria incidence of about 328.2 between 1998-1999 and 2004-2005 (Gerritsen *et al.*, 2008). In 2000, the municipality has experienced floods brought by the tropical cyclone *Eline* which have dramatically increased malaria cases in Limpopo province (Reason and Keibel, 2004). Malaria in Limpopo is seasonal and it is, therefore, crucial to use seasonal data covering the entire study area to map the occurrence of malaria in the VDM.



**Figure 5.1.** The study area in the northern part of Limpopo. Each coloured polygon represents individual villages under consideration ( $n = 28$ ).

### 5.2.2. Epidemiological data

This chapter used secondary data acquired from the malaria information systems (MIS) of the South African National Department of Health that were developed and maintained by the malaria control programme (MCP). Ethical approval for this chapter was obtained from the faculty of Natural and Agricultural Science Ethical Committee at the University of Pretoria (No. 180000076). The epidemiological data used for the study was obtained from the South African National Department of Health's Malaria Information Systems (MIS). The dataset comprises of the presence cases of malaria agent (*P. falciparum*) in rural villages in Vhembe District,

from 1998-2006. It was obtained passively from patients who tested positive for *P. falciparum* in health centres around the selected villages. A total of 28 presence locations at village level was recorded for the year 2005. In the current study the absence data was originally unavailable since the surveys conducted were designed to report on positive malaria cases by health-care workers. The absence (pseudo-absence) points were generated and were two times (2x) the number of presence points which formed part of a standard dataset to be used for modelling.

Generally the MIS dataset comprises of additional records such as the locality, facility names where malaria tests were conducted and the source of infection. Because the Vhembe District Municipality forms the border between South Africa and Zimbabwe/Botswana, some of the cases reported in this area are imported from Mozambique, Zimbabwe and Botswana. For the purpose of this study, all the imported cases outside of the VDM were filtered so that an understanding of local environmental factors can be derived from conditions existing within the municipality.

### **5.2.3. Remote sensing data**

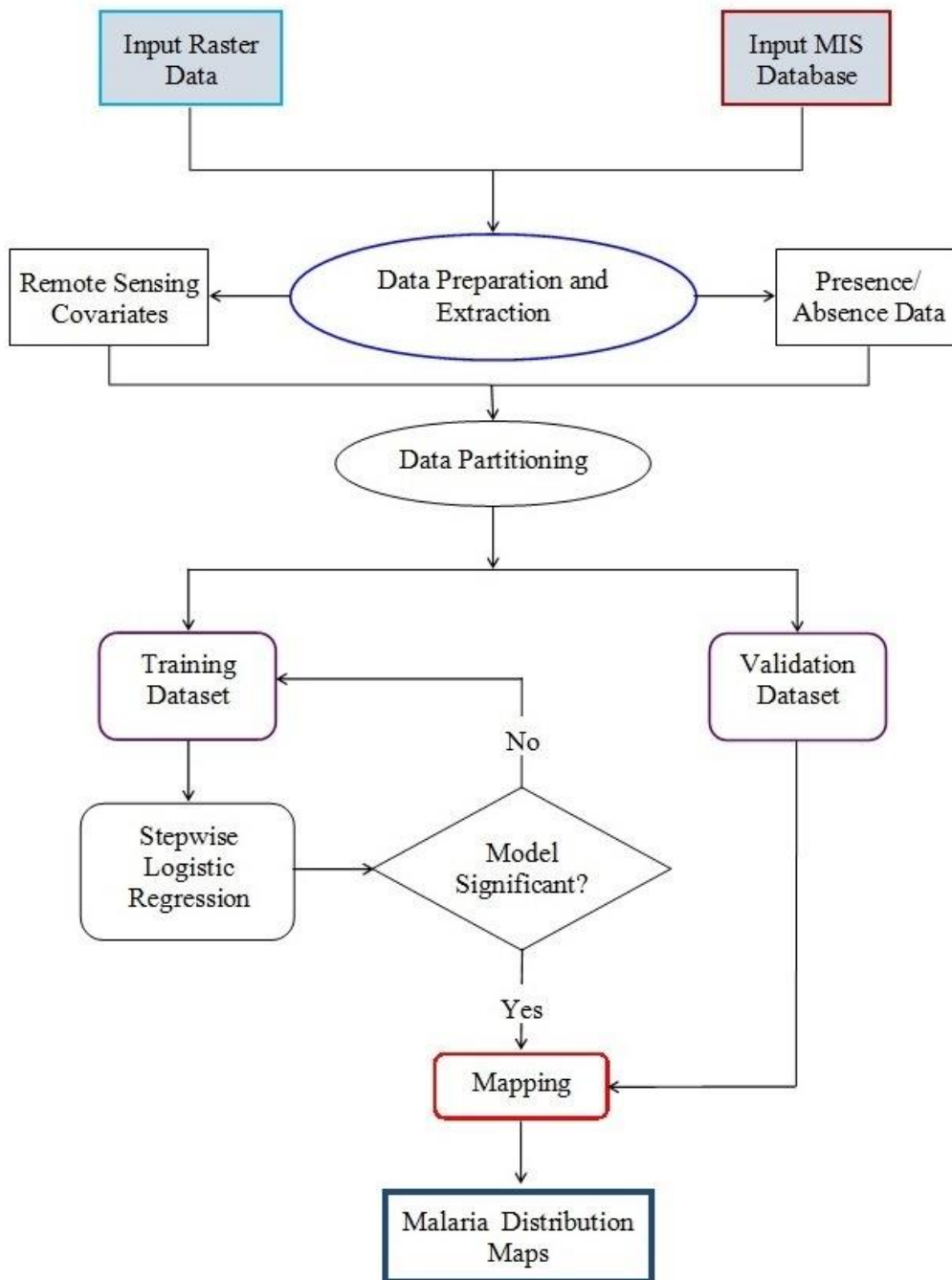
The Landsat 5 Thematic Mapper (TM) imagery acquired in spring of 2005 were used for the study. Three multispectral images were acquired from United States Geological Surveys website (<http://earthexplorer.usgs.gov/>). Landsat 5 comprises of 7 multispectral bands, in the visible to thermal infrared region (0.45-12.5  $\mu\text{m}$ ) at 30 meters spatial resolution. Images were acquired on path and rows 169/76, 170/75 and 170/76 (12 September 2005), and for paths/rows 170/75 and 170/76 (19 September 2005). This period corresponds to the rising malaria incidences in the study area (Gerritsen *et al.*, 2008). The multi-spectral bands of Landsat TM are commonly used for vegetation, bathymetric, and soil moisture mapping. Vegetation is one of the environmental factors, depending on climatic evolutions, that influences malaria vector behaviour directly or indirectly (Gomez-Elipe *et al.*, 2007). Therefore computing vegetation indices that are sensitive to changes in vegetation greenness could enhance the understanding of malaria patterns. Various environmental covariates were generated from Landsat TM which relate to vegetation biogeophysical/chemical properties and moisture. Perhaps one of the most commonly used satellite-derived indices is the normalized difference vegetation index (NDVI) which is primarily used as an indicator of vegetation greenness and biomass (Jackson *et al.*, 1983). In addition to Landsat data, a 30 m

digital elevation model (DEM) data from Shuttle Radar Topography Mission (SRTM) was used to derive the aspect of individual presence-absence points.

#### **5.2.4. Data pre-processing**

Figure 5.2 shows the workflow adopted in this chapter. The processing of epidemiological data was done in Microsoft Excel spreadsheet. Firstly, the presence dataset was prepared according to the number of villages with records of *P. falciparum* presence. A total of 28 villages have been extracted from the MIS dataset, which contain recorded malaria cases. The original MIS dataset contained the name of the village, local municipality, health centres, locality, death status, age and sex of the infected which were not employed for the study.

The pre-processing of Landsat TM involved four (4) stages: (i) band merging, (ii) atmospheric correction, (iii) image mosaicking, and (iv) study area subsetting. Firstly, six Landsat TM spectral bands were merged to derive a 6-band multi-temporal image composite, excluding thermal bands. This process was applied on individual scenes with similar metadata file definition (.mtl). The subsequent merging of spectral bands was carried out in Quantum Geographic Information System (QGIS) software (QGIS Development Team, 2016). Secondly, in order to minimize the influence of atmosphere on data analysis, all of the merged multi-spectral images were subjected to atmospheric correction in ENVI version 4.7 through the Quick Atmospheric Correction (QUAC) module (Exelis Visual Information Solutions, 2016). The QUAC approach is based on the empirical finding that the mean spectrum of a collection of diverse material spectra, such as the end-member spectra in a scene, is essentially invariant from scene to scene (Bernstein *et al.*, 2012). Thirdly, the atmospherically corrected images were mosaicked in ENVI 4.7 to derive a single large image that covers the entire study area. In total three (3) images were stacked to form part of a larger multi-spectral image. And finally, the final image of the Vhembe District Municipality (VDM) was extracted from the image mosaic by use of corresponding municipal shapefile (.shp). The spatial mask was created for the high altitude terrain along the Soutpansberg Mountain, which comprises of forest vegetation that does inhibit malaria transmission through complete shading effect (Kamau *et al.*, 2006).

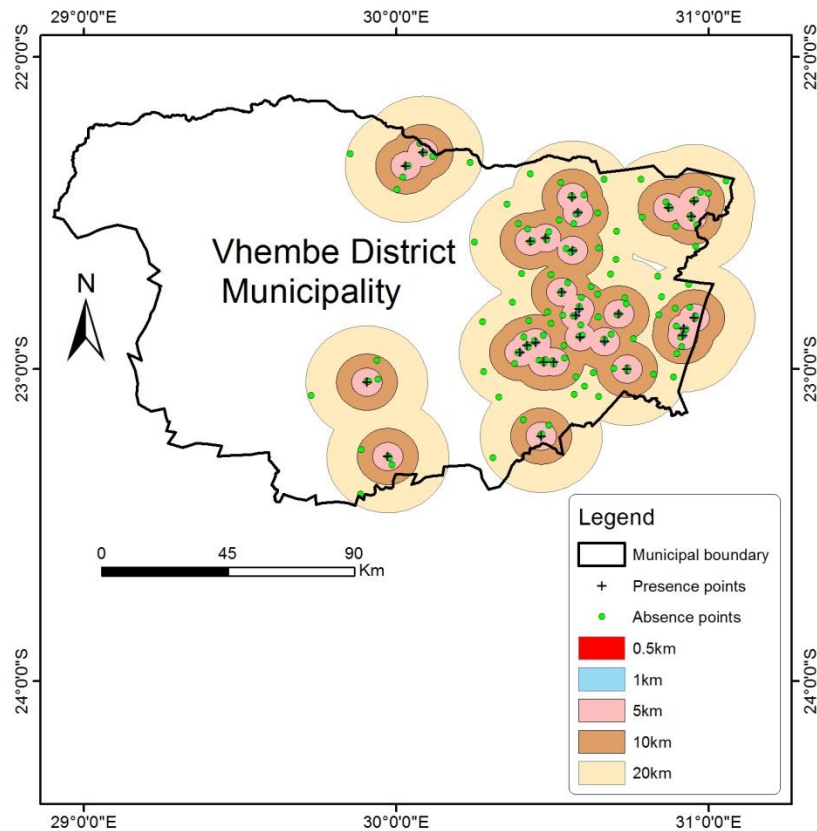


**Figure 5.2.** Overview of epidemiological and remote sensing datasets and methods used for the study



### 5.2.5. Data processing and analysis

The pseudo-absences were generated for five buffer distances from the known presence locations. To test for the utility of Landsat 5 data for malaria mapping, the buffer distances were set at 0.5 km, 1 km, 5 km, 10 km, and 20 km (Figure 5.3). The choice of buffer distances was dictated by the proximity of one village to the neighbouring villages other so as to avoid possible duplication of occurrence points. The generated pseudo-absences formed part of the standard dataset used for modelling. In total, 56 pseudo-absences ( $n = 56$ ) were derived, making presence-absence ratio of 1:2. The standard *P. falciparum* points were equal to 84 ( $N = 84$ ). The dataset was geo-referenced using World Geodetic System (WGS-84) and exported in GIS software in order to allocate individual location ID in the dataset.



**Figure 5.3.** Representation of different buffer distances across 28 villages in Vhembe District Municipality. Buffer distances were ranging from 0.5km to 20 km.

### 5.2.6. Remote sensing data analysis

A number of remotely-sensed indices were computed in order to assess their usefulness for mapping malaria distribution in the study area. These indices were computed based on the environmental factors that are known to influence breeding patterns and survival of malaria

vectors. Environmental factors such as vegetation moisture content, the greenness and daily temperatures of an area are known to have an impact on malaria spread, vector reproductive rates, and pathogen incubation period (Wayant *et al.*, 2010; Baeza *et al.*, 2011). It was upon this premise that the normalized difference vegetation index (NDVI), modified normalized difference water indices (MNDWI<sub>1</sub> and MNDWI<sub>2</sub>), green index (GI), and soil adjusted vegetation index (SAVI) were computed from Landsat 5 data and tested for the study. The common NDVI is highly susceptible to errors over canopy and soil background in VDM, especially in September month where large parts of the land have low vegetation cover. Additional to these indices, a quasi-yellowness index (p-YI) was derived due to its possible relationship with the habitats of *An. arabiensis* in the study area (Adams *et al.*, 1999). The yellowness index was first introduced to estimate chlorosis, although its application may not be limited to plant health and stress analysis (Malahlela *et al.*, 2014). Its inverse correlation to NDVI could potentially shed light on patterns of malaria occurrence in semi-arid environment such as the VDM, considering that high malaria prevalence is usually observed during high NDVI season, often characterized by high daily temperatures, rainfall, humidity and high chlorophyll composition in plants (Komen *et al.*, 2018). Both NDVI and yellowness index are related to chlorophyll concentration and thus p-YI was incorporated in the study of malaria. Other remotely sensed indices were designed for the study based on the relationship between spectral bands and moisture. These indices are named moisture indices ( $a_1$  and  $a_2$ ) as shown in table 5.2. These indices were computed by considering reflectances in shortwave infrared spectral region (1.55-2.35  $\mu\text{m}$ ) to malaria mapping, which may have potential confounding effect to MNDWI<sub>1</sub>. The shortwave infrared (SWIR) reflectance generally decreases as water content in the leaves increases at 1-3 $\mu\text{m}$  (Gao, 1996; Hunt and Rock 1989) and in this chapter average reflectance and the spectral difference between NIR and SWIR were explored. The shortwave infrared bands are sensitive to soil moisture, changes in vegetation moisture and water bodies which are potential habitats and breeding sites for *An. arabiensis* species. (Bowman, 1989; Tucker, 1980).

In addition, the aspect (direction to which the slope faces) was derived from Advanced Spaceborne Thermal Emission Reflection Radiometer (ASTER)'s digital elevation model (DEM), that has similar spatial resolution as Landsat TM of the study area.

**Table 5.1:** Selected remote sensing indices that were employed for the study. All indices are derived from Landsat TM

Index	Formulation	Reference
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{(\rho_{NIR} - \rho_{Red})}{(\rho_{NIR} + \rho_{Red})}$	Jackson <i>et al.</i> (1983)
Modified Normalized Difference Water Index (MNDWI <sub>1</sub> )	$MNDWI_1 = \frac{(\rho_{NIR} - \rho_{SWIR})}{(\rho_{NIR} + \rho_{SWIR})}$	Ceccato <i>et al.</i> (2001)
Modified Normalized Difference Water Index (MNDWI <sub>2</sub> )	$MNDWI_2 = \frac{(\rho_{Green} - \rho_{MIR})}{(\rho_{Green} + \rho_{MIR})}$	Xu (2006)
Moisture index ( <i>a</i> <sub>1</sub> )	$a_1 = \left( \frac{\rho_{SWIR} + \rho_{MIR}}{2} \right)$	<i>In this chapter</i>
Moisture index ( <i>a</i> <sub>2</sub> )	$a_2 = \rho_{NIR} - \rho_{SWIR}$	<i>In this chapter</i>
Soil adjusted vegetation index	$SAVI = \frac{(\rho_{NIR} - \rho_{Red})}{(\rho_{NIR} + \rho_{Red} + L)} (1 + L)$	Huete (1988)
Green index	$GI = \frac{\rho_{NIR}}{\rho_{Green}}$	Gitelson (2003)
Quasi-Yellowness Index	$p-YI = \frac{\rho_{Yellow} - (2 * \rho_{Green}) + \rho_{Red}}{\rho_{Blue}^2}$	Adams <i>et al.</i> (1999)

### 5.2.7. Statistical analysis

Statistical analysis was performed on a full standard dataset comprising of presence and pseudo-absence locations together with the environmental covariates derived from Landsat TM data acquired in spring 2005. To calibrate the statistical model, 60% ( $n = 50$ ) of the data was used for training, while independent 40% ( $n = 34$ ) was used for validating the predictive model. The stepwise logistic regression (SLR) model has been applied to training dataset derived at five (5) buffer distances in R software (R Development Core Team, Vienna), using a glm2 package. Both backward and forward SLR were used in order to select covariates with

high relative importance using relaimpo package, so as to avoid multi-collinearity issues and model over-fit (Collet, 1991). The automated SLR is one of the common statistical methods for public health, by relating the remotely sensed data with the disease distribution such as malaria (Adimi *et al.*, 2010; de Oliveira *et al.*, 2013). The automated procedure for variable selection has advantages in that it reduces computation time and tedious manual modelling, especially in large, complex candidate models (Ripley, 2003; Calcagno and de Mazancourt, 2010). The choice of the stepwise logistic regression model was dictated by the binary nature of the response variable (presence/absence), its simplicity for embedding in GIS environment (Yang *et al.*, 2006) and its popularity amongst all other predictive models. The logistic regression is given by the equation (5.1):

$$y_i = \frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij}\right)\right]} \quad (5.1)$$

where  $y_i$  is the probability of malaria distribution (1 or 0),  $x_i$  is the environmental covariate at the  $j$ th location,  $\beta_j$  is the coefficient of  $x_n$ ,  $\beta_0$  is an intercept, and  $\exp$  is the exponential function of the regression. The malaria distribution maps were derived from final selected models, with lowest Aikake's Information Criterion (AIC) as these represented best fit. In addition, probability maps with threshold values of greater than 0.7 were produced, since model performance at his threshold is deemed a good model (Baldwin, 2009). The bootstrap resampling was performed on the independent validation dataset to assess the robustness of the regression. The validation dataset was bootstrapped with the replacement for  $n = 10\ 000$  time using the boot package in R. The coefficient of variation (CV) was used as a measure of variability of the pseudo-absences of validation dataset.

The model deviance ( $D^2$ ), which is an analogy of  $R^2$ , was used to determine the percentage of variability explained by the remote sensing covariates. The  $D^2$  is given by the equation (5.2):

$$D^2 = 1 - \left(\frac{\rho\sigma_1}{\tau\rho w}\right) \quad (5.2)$$

where  $\rho\sigma_1$  is the residual deviance, and  $\tau\rho w$  is the null deviance. A good model is the one with low AIC and high  $D^2$ . In addition, the F-test was performed to measure the variances between measured and predicted presence/absence locations of *P. falciparum*. In order to assess the validity of the model, the overall classification accuracy was determined from the independent validation dataset. The overall accuracy (OA) is the number of correctly classified cases (presence/absences) to the total number of cases in the dataset. Classification accuracy was

done on probability threshold values of 0.5-1.0. The reason for leaving out lower probability threshold values was that a model with probability values of < 0.5 is considered a failed model (Baldwin, 2009).

### 5.3. Results

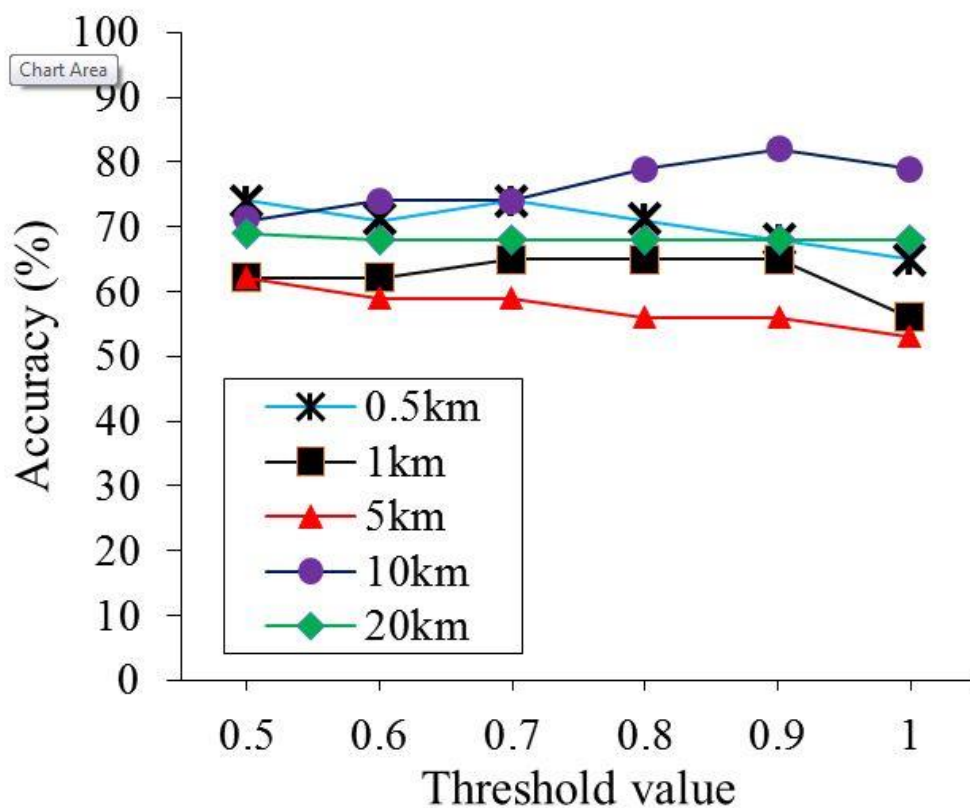
The results of the study are presented in table 5.2. The table showed that SAVI was the most significant remotely-derived covariate in predicting the distribution of malaria in VDM, Limpopo.

**Table 5.2:** Results of the logistic regression along 5 buffer distances from the known *P. falciparum* locations.

Buffer distance (km)	Covariates	Estimates	$\rho$ value ( $>  z $ )	Significance level
0.5	$\beta_0$	-36.370	0.008	**
	NDVI	108.100	0.043	*
	SAVI	124.340	0.027	*
	$a_1$	-46.220	0.012	*
	$a_2$	63.250	0.017	*
1.0	$\beta_0$	0.629	0.755	
	SAVI	14.882	0.041	*
	NDWI <sub>1</sub>	45.984	0.078	
	$a_2$	87.529	0.094	
	p-YI	-7.241	0.013	*
5.0	$\beta_0$	5.341	0.001	***
	SAVI	13.659	0.012	*
	Aspect	0.007	0.061	
10.0	$\beta_0$	-43.551	0.053	
	NDVI	131.350	0.021	*
	SAVI	159.840	0.013	*
	$a_1$	-58.851	0.057	
	$a_2$	181.070	0.096	
	NDWI <sub>1</sub>	89.790	0.049	*
	NDWI <sub>2</sub>	-36.030	0.016	*
20.0	$\beta_0$	14.904	0.006	**
	NDVI	84.780	0.009	**
	SAVI	90.310	0.008	**
	$a_1$	-20.683	0.017	*

Significance codes: \*\*\* (0.001), \*\* (0.01), \* (0.05)

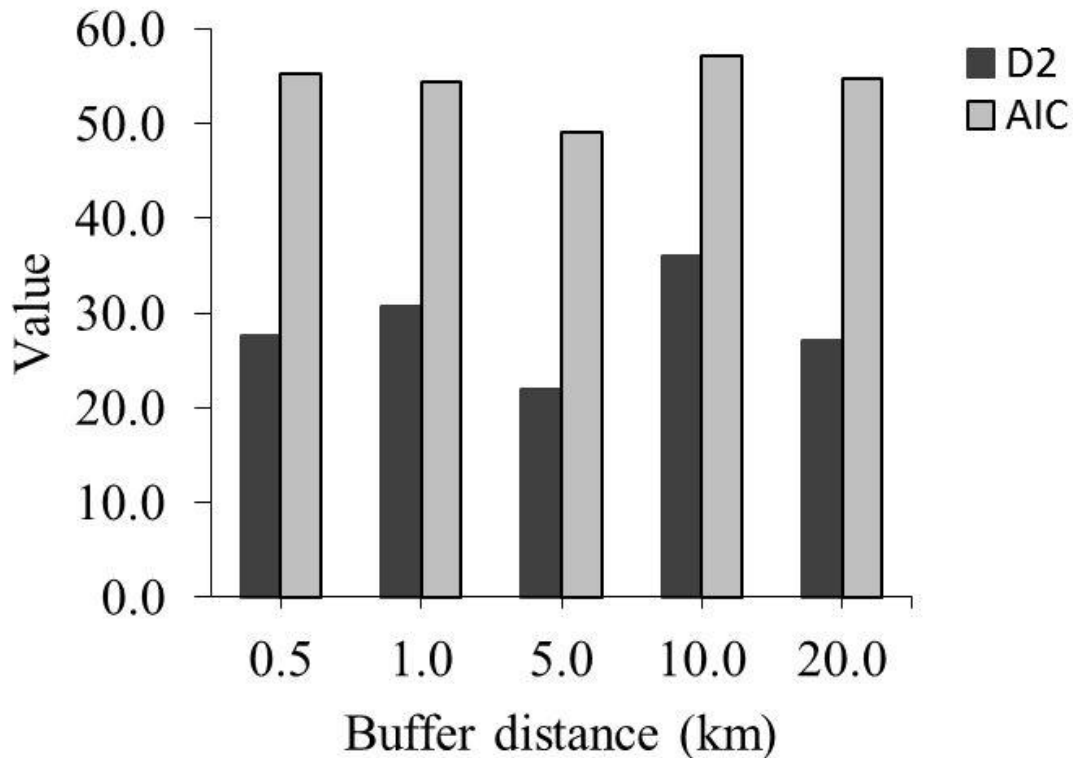
In all the derived models, the SAVI showed a significant positive correlation with the distribution of *P. falciparum* (malaria) ( $\rho < 0.05$ ). The results also show that NDVI is significantly correlated with the distribution of *P. falciparum* at buffer distances of 0.5km, 10km, and 20km from the presence locations. The introduced indices which are sensitive to moisture changes ( $a_1$  index and  $a_2$  index) have shown to be negatively correlated with the distribution of malaria pathogen, although they are mostly significant at buffer distances of 0.5 km and 20 km. Predicting the distribution of *P. falciparum* at 10 km distance yielded the highest classification accuracy of 82% at a threshold of 0.9 (Figure 5.4), while at 5 km low classification accuracy (54%) was found at the threshold value of  $\rho = 1.0$ .



**Figure 5.4.** Logistic regression performance across threshold values of 0.5-1.0 as applied on buffer distances selected for the study area using validation dataset ( $n = 34$ ).

In the current study the highest variation explained by the predictive model was found to be 36% ( $D^2 = 0.36$ ;  $AIC = 57.07$ ) 10 km away from the known presence locations, while the lowest explained variation (27%) was found at buffer distance of 20 km away from the known

presence location ( $D^2 = 0.27$ ;  $AIC = 54.73$ ). Figure 4.5 shows the results of stepwise logistic regression as applied on distances 0.5 km-20 km from the known *P. falciparum* presence location. The validation dataset was bootstrapped and yielded the CV of 0.11, indicating the small variation of the pseudo-absences from the calibration dataset.

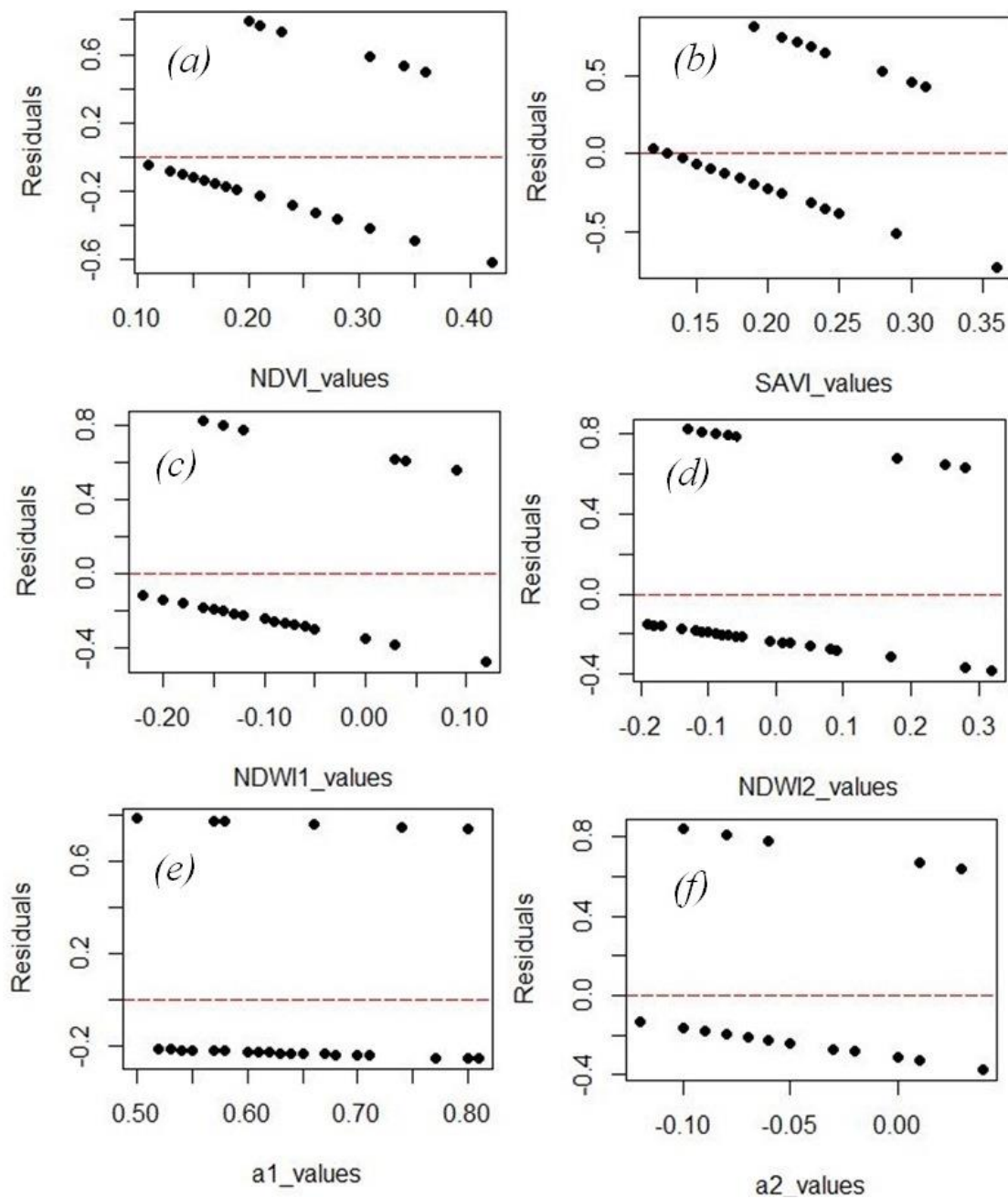


**Figure 5.5.** The  $D^2$  and AIC of logistic regression applied on buffer distances in VDM.

The F-test performed between the measured and predicted presence/absence data yielded a statistical  $p$  value of 0.081, which is greater than the  $p = 0.05$ . This result indicates that there was no significant difference between the observed and predicted malaria occurrence in the rural villages of VDM. The SLR model shows that the NDVI, SAVI,  $NDWI_1$  and  $NDWI_2$  were statistical significant variables and thus improved the prediction of malaria occurrence when compared to a 20 km p/a model. Ideally, a model that explains greater than 50% ( $D^2/R^2 > 0.5$ ) of the variations in the p/a occurrence of species is considered a relatively good model (Lopatin *et al.*, 2017). Although all the models tested for the current study yielded a  $D^2$  of less than 40%, the ultimate model used to create predictive maps was significant at  $p=0.033$ . The variable residual plots (Figure 5.5) show the residuals, with NDVI, SAVI, and  $NDWI_1$  exhibiting lower residual deviance. However, randomly collecting the pseudo-absences at various spatial



scales and at highly heterogeneous village areas such as in VDM is very challenging for adequately selecting environmental variables that best explain the distribution of *An. arabiensis* species. Ecological systems are complex, therefore multi-temporal data about the interactions between environmental covariates and malaria distribution are required to untangle such complexities.



**Figure 5.6.** Residual plots of environmental covariates used for mapping malaria distribution in VDM.

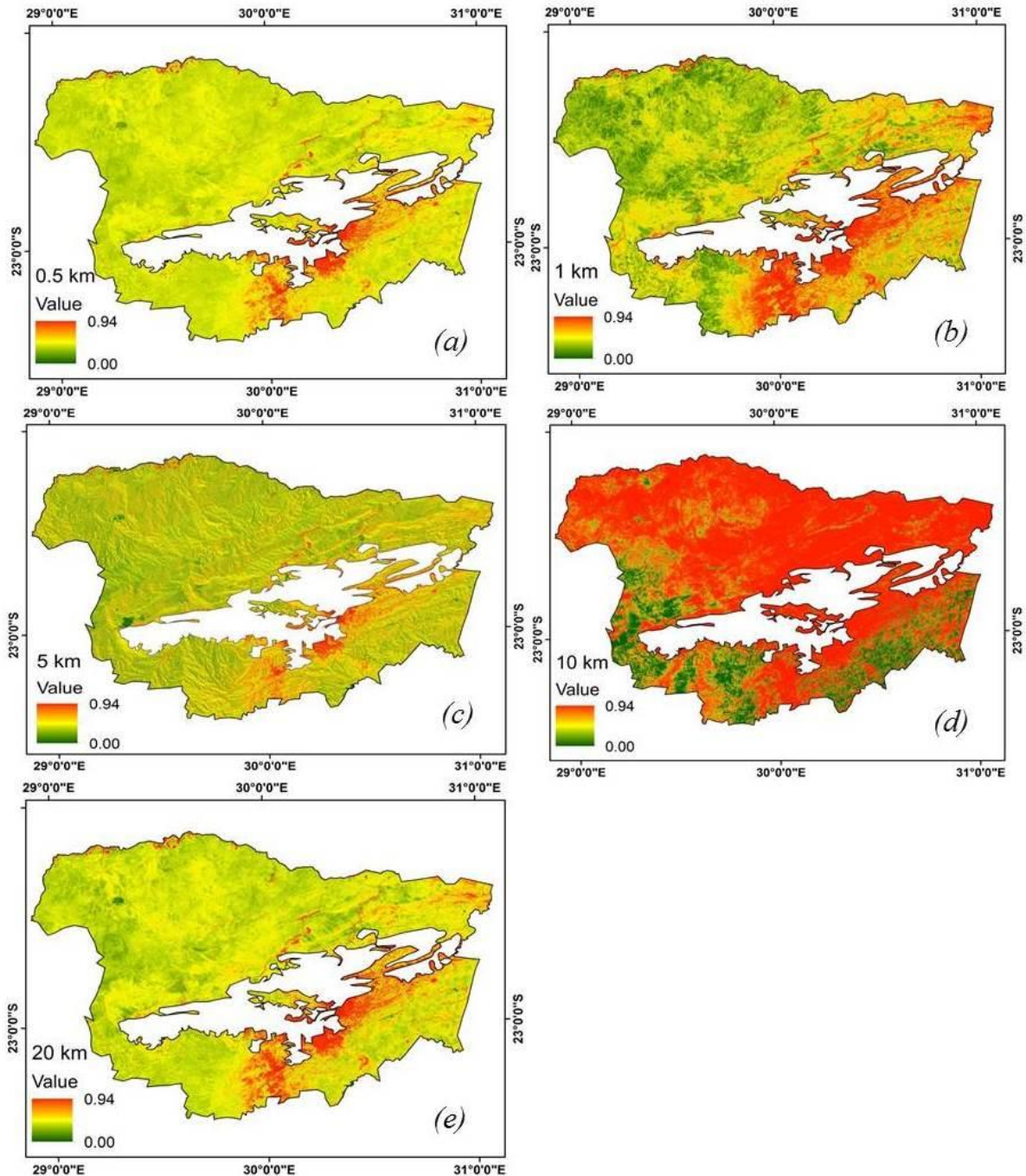
### 5.3.1. Predictive maps

The results of the predicted probability maps are shown in figures 5.7 (a-e) below. From these results it becomes apparent that the predicted malaria distribution exhibits spatial heterogeneity across 5 buffer distances, which may be attributed to landscape configuration and environmental factors used for malaria modelling. The model calibrated from pseudo-absence dataset within 0.5 km radius produced a much narrower trend in malaria distribution (Figure 5.7 a). In areas at the foot of Soutpansberg Mountain the model has predicted a high probability of malaria distribution due to apparently high ground cover, comprising of moist, riparian vegetation. Conversely, the wide distribution of malaria pathogen is recognized when pseudo-absences within 10 km are used (Figure 5.7 d). This pattern, which is mapped at high classification accuracy, indicates that cases of malaria could potentially be detected in areas that were previously considered unsuitable for the survival of *P. falciparum* pathogen and its vector.

Additional to malaria probability maps that range from 0 (less likely) to 1 (more likely), spatial heterogeneity of malaria distribution was examined at threshold greater than 0.7. Figure 5.8 (a-e) shows the results of this threshold applied onto the predictive maps. In general, high probability of malaria is noticeable in three areas of concern: (i) at the settlements at the foot of Soutpansberg Mountain, (ii) along the riverine areas and (iii) closer to low-lying irrigated fields.

## 5.4. Discussions

The aim of this chapter was to assess the feasibility of Landsat-derived environmental covariates for predicting malaria distribution in rural landscape of Vhembe District Municipality in South Africa, across different buffer distances. Malaria pseudo-absences with ratio of 1:2 (presence-absence) were randomly generated and constrained within buffer distances of 0.5 km, 1 km, 5 km, 10 km, and 20 km. In epidemiological studies, the vast majority of species occurrence datasets are subject to spatial bias, such as sampling near roads which often leads to some areas surveyed more frequently than others. The selection of different buffer distances and randomly generated pseudo-absences ensures that spatial environmental bias is minimized and improves the model performance and robustness (Moyes *et al.*, 2016; Phillips *et al.*, 2006). In this chapter the results have shown that intermediate buffer distance (10 km) yielded the highest classification accuracy (82%) at threshold of 0.9.



**Figure 5.7** Predicted potential geographic distribution of malaria produced through logistic regression and Landsat-derived environmental covariates. Maps produced at buffer distances of 0.5 km (a), 1km (b), 5 km (c), 10 km (d) and 20 km (e).

This may be an indication that the generated pseudo-absences at this distance represented areas of low suitability for *An. arabiensis* species occurrence, which is the dominant *P.*

*falciparum* vector in VDM area. Additionally, higher classification accuracy at 10 km buffer distance may show the similarity between pseudo-absences and ‘true absences’ where *P. falciparum* vector is likely to occur ( $CV = 0.11$ ,  $\rho = 0.3$ ). These findings demonstrate the potential of medium resolution satellite data to predict malaria distribution at local levels (high spatial resolution), as most studies focused on regional and global patterns. Consequently, areas with the highest likelihood of malaria occurrence are located around the green vegetated environments which serve as refuges for malaria vector (Ricotta *et al.*, 2014). The lowest classification results attained at 5 km from the presence location may indicate the probability of pseudo-absences to have fallen on the unidentified presence location, particularly consideration the random nature of their derivation (Chefaoui and Lobo, 2007). Whether this effect is peculiar to a medium spatial resolution dataset (30m), such as Landsat TM/ASTER or high spatial resolution datasets (e.g. SPOT), is a matter that requires further research.

#### **5.4.1. Remote sensing environmental covariates**

In all models, SAVI exhibited a statistically significant pattern as a remote sensing-derived covariate at  $\rho < 0.05$ . The findings from the current study differ from those obtained by Jacob *et al.* (2007) who concluded that NDVI, SAVI and atmospherically-resistant vegetation index (ARVI) were not related to ecological conditions necessary for *An. arabiensis* habitat suitability. This difference could be attributed to methodology used in this chapter, where the pseudo-absences were utilized while Jacob *et al.* (2007) opted to use no pseudo-absences generated at 30 m spatial resolution. The fact that Jacobs *et al.* (2007) did not subject the image (Quickbird) to atmospheric correction process which reduces the influence of atmospheric noise, might have contributed to the subsequent correlations between vegetation indices and *Anopheles* mosquitoes habitats. In addition, there was no apparent description of threshold values defining vegetation range for indices used (e.g. NDVI), except for the binary land cover classification of paddy vegetation which might have affected variable contribution in modeling. In contrast, the current study is one of the first studies in South African semi-arid environment that assessed correlation of remote sensing covariates to *P. falciparum* distribution at spatial resolution higher than 50 meters. Some authors have found NDVI to be the strongly correlated covariate than many other indices in malaria studies elsewhere in Africa (Machault *et al.*, 2010). In contrast, the current chapter has shown that the NDVI correlation is environment dependent and therefore in semi-arid environments SAVI, which takes into account effect of soil background, performs higher than NDVI. Moreover, NDVI is highly susceptible to errors over canopy and soil background in VDM, especially in September month



where large parts of land have low vegetation cover. In addition, it has been documented that NDVI may suffer from signal saturation especially when used in dense vegetation (Malahlela *et al.*, 2014). The significant positive correlation of SAVI with malaria distribution at 0.5-20 km buffer distances shed light on environmental conditions suitable for survival of malaria vector (*An. arabiensis*). It is generally known that SAVI values increase with an increase in the vegetation greenness and biomass (Huete, 1988; Araujo *et al.*, 2000). The abundance of healthy vegetation provides mosquitoes with resting sites and refuges (Ricotta *et al.*, 2014), thereby intensifying foci for malaria transmission.

One of the notable correlations is between malaria occurrence and spectral indices from shortwave-infrared bands ( $a_1$  index and  $a_2$  index). Both the indices were firstly tested based on the assumption that shortwave infrared band (SWIR) is inversely correlated to vegetation water content, and therefore, with a probability of malaria distribution. Although each plant has its own relationship with chlorophyll content and vegetation water content, the first moisture index ( $a_1$ ) has explained the general relationship of vegetation water content and a probability of malaria occurrence in the study area (Ceccato *et al.*, 2002). The second moisture index ( $a_2$ ) was a measure of vegetation water content that showed a positive correlation with malaria pathogen, although the statistical significance was only found at buffer distance of 0.5 km ( $\rho < 0.05$ ; AIC = 55.33). However, the sensitivity of this index in other environments is subject for further research. This chapter has shown that in instances where NDVI shows no significant association with malaria distribution or risk, other indices such as SAVI and  $a_2$  can be used instead. On the other hand the MNDWI<sub>1</sub> has shown to be significantly positively correlated to the malaria distribution at buffer radii of 0.5 km ( $\rho < 0.05$ ; AIC = 55.33) and 10 km ( $\rho < 0.05$ ; AIC = 57.07).

The correlations of MNDWI<sub>1</sub> at 0.5 km and 10 km buffer distances show that availability of water bodies is crucial for survival adaptations of *An. arabiensis* mosquitoes at the study area. The correlation found in this study between MNDWI<sub>1</sub> and malaria distribution has also been established elsewhere in Africa by Dambach *et al.* (2012). Water bodies serve as breeding sites for mosquitoes, although the preferences in terms of the size, compactness, depth, temperature and quality differ from one *Anopheline* species to another (Zhou *et al.*, 2012). The correlation between *P. falciparum* pathogen with NDVI, SAVI, moisture indices ( $a_1$  and  $a_2$ ) and MNDWI<sub>1</sub>, may serve as an indication that malaria distribution is correlated to response of vegetation to rainfall and temperatures. This is particularly true in that vegetation index NDVI is known to be a surrogate for rainfall (Mabaso *et al.*, 2006). Although not statistically

significant the p-YI has exhibited negative correlation with the distribution of malaria in the study area using 5 km model. This is an indication that indeed green vegetation intensifies malaria transmission in that p-YI is negatively correlated to vegetation greenness, and therefore malaria distribution. The use of these indices can form part of the malaria early warning systems in support of eradication efforts. The improved spectral and radiometric resolution of Landsat satellite (currently Landsat 8) could be used to detect the breeding and questing sites for *An. arabiensis* mosquitoes at village level at higher accuracies, thus reducing costs associated with manual surveying of environments surrounding them.

The findings also show that the orientation of slope (aspect) towards the north results in an increased probability of malaria occurrence, particularly at buffer distance of 5 km from the known presence locations. This is especially true, in that malaria vectors prefer warm, moist habitats and therefore, the inclination of the slope towards the north create optimal conditions for *An. arabiensis* habitation. Most of the villages are situated on the north-eastern slopes whose surrounding vegetation provide resting and refuges for mosquitoes. It is generally known that in the southern hemisphere north-facing slopes are warmer than south-facing slopes (Adams, 2010), and therefore integration of remote sensing data and temperature/rainfall data could enhance insight into malaria control and eradication. The information can be used by the policy-makers and the health-care professional to distribute the limited financial resources to the areas that are highly affected by malaria. In addition, further financial investments have to be allocated to the areas that were previously known to be malaria, such as the western and the north-western part of the VDM. This study shows that climate change may alter the traditional habitable environments for malaria by extending the plasticity of *An. arabiensis* across the semi-arid environments.

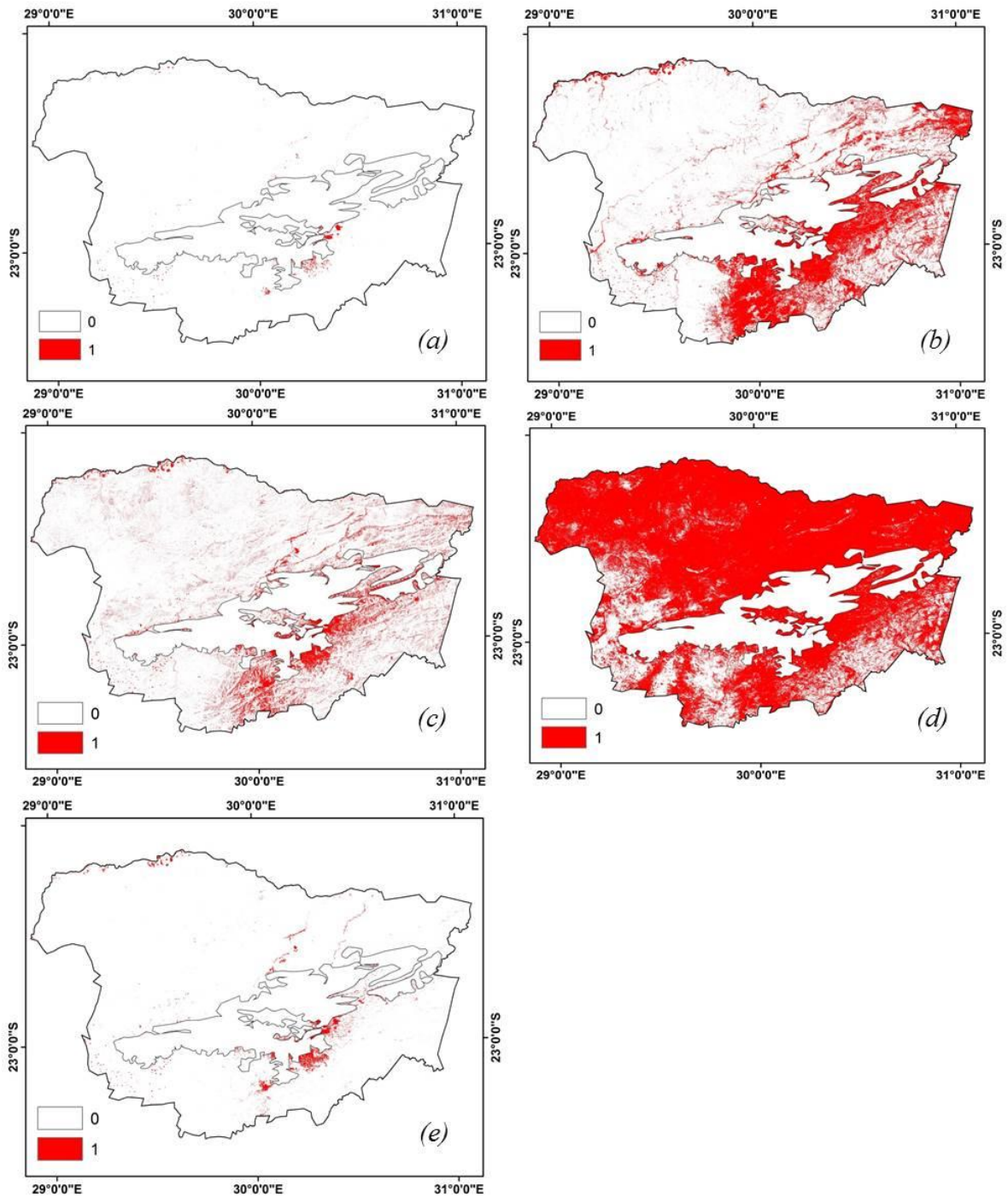
#### **5.4.2. Predictive maps**

The pattern of malaria distribution is highlighted by the probability maps in figure 5.7. These figures show varying degrees of the probability of malaria occurrence across the VDM. In comparison, the map produced from pseudo-absences that were derived from 0.5 km (Figure 4.7 a) shows a rather narrow spatial pattern than both the 1 km and 10 km map. This map is produced from low accuracy model, in which probability of malaria occurrence stretches up to the traditionally malaria-free southern part of the VDM. In figure 5.7 (b) high probability of malaria occurrence stretches from east (Mutale and Thulamela local municipalities) to the western part of the VDM, with high occurrences predicted at the low-lying areas of the Thulamela local municipality. In contrast to the 0.5 km and 1 km predictive maps, the 5 km

buffer distance predictive map relates the high probability of malaria occurrence to the aspect or slope orientation. In this map, the villages located at north-facing slopes (0.0-67.5°) and the east-facing slopes (67.5-112.5°) exhibit high likelihood of malaria transmission than both the south and west-facing villages.

Slope orientation plays a major role in the distribution of floral composition over a long period (Bennie *et al.*, 2006). The map derived from 10 km pseudo-absences radius showed high probability of malaria occurrence in the eastern part of the VDM, where Mutale and Thulamela municipalities are located, which are areas known for high malaria transmission in the Limpopo province of South Africa (Khosa *et al.*, 2013; NICD-NHLS, 2017). In all the predictive maps, it appears that very low probabilities of malaria occurrence are found within the Makhado local municipality in the south western part of Soutpansberg Mountain.





**Figure 5.8** Predicted spatial distribution of malaria in VDM at a probability threshold of  $\geq 0.7$ . Red colour depicts areas of predicted *P. falciparum* presence, while white represents predicted *P. falciparum* absence and Soutpansberg Mountain mask.

### 5.4.3. Limitations and recommendations

Although this chapter has successfully utilized remotely sensed data for mapping malaria distribution, it had its own limitations. One of the limitations emanates from the routine data collection from the national malaria control centre. The data collection system was mainly passive, with more room left for possible under-diagnosis or under-reporting. This could be averted by continuous sampling and the improvement of the methods used for sampling, which often requires considerable monetary investments. There are more than 28 villages in VDM, which could otherwise be included in the analysis, but were excluded due to lack of data. The inclusion of cases from other villages, although they may be few in number, could have potentially increased the number of presence locations, and therefore accuracy as a large number of cases (presence-absence) and iterations have a positive impact on statistical  $\rho$  value (Berkson, 1938). Another limitation of the study is the manner in which the random pseudo-absences were generated. It is recommended that the same geographical bias adopted for presence points be used even for pseudo-absences (Phillips *et al.*, 2009). The pseudo-absences' geographic location could have been the unidentified presence locations for *P. falciparum* due to the sampling protocol adopted in this chapter. Malaria transmission at the study area is largely influenced by environmental factors including temperatures and rainfall (Komen *et al.*, 2015; Komen, 2017). In addition, Ikeda *et al.* (2017) have concluded that incidences of malaria in Limpopo province are positively correlated to the lag in local and climatic systems (e.g. rainfall) that occur in neighbouring countries. The use of temperature and rainfall data could essentially assist in the improvement of model prediction thus enhancing the overall classification  $D^2$ . Thus, assessing the impact of climate change on malaria transmission requires consideration of not only annual mean temperature changes, but more importantly, the extent of temperature and rainfall interannual variability (Zhou *et al.*, 2004). Unfortunately, the use of optical Landsat data is largely dependent on cloud-free atmosphere which may serve as a limitation to time-series analysis. The integration of active remote sensing data with optical could increase the temporal and spatial coverage of malaria endemic areas thus aid in mapping the disease occurrence during the periods of high rainfalls.

### 5.5. Conclusions

From the results, it can be concluded that remotely sensed data can be used for mapping malaria in a semi-arid environment. The study has also shown that deriving the pseudo-absences at the intermediate distances (approximately 10 km) from the known presence location yields high classification accuracies than drawing pseudo-absences at very far or very

near distances. The remotely sensed variables such as the SAVI and NDVI serve as good indicators for the environmental conditions that encourage *An. arabiensis* reproduction, questing and malaria transmission rates. If South Africa is to eradicate malaria by the year 2020, there should be intensified efforts towards early detection of the environmental/local conditions commonly associated with malaria spread, especially in rural Vhembe District, which has high malaria rates. The use of the latest Landsat data coupled with ancillary epidemiological and climatic data should form the integral part of the malaria early-warning system due to the frequency of data satellite data acquisition (16 days cycle) and the biological/ecological nature of the *P. falciparum* vector. The findings from this chapter serve as baseline information for developing methodology necessary to detect and model malaria pathogen, vector, and habitat preference through the use of earth observation techniques.

In summary, this chapter successfully demonstrates how Landsat data helps in mapping the distribution of *P. falciparum* in the malaria endemic environment. In this study, it was found that malaria occurrence is highly correlated to villages with the surrounding green vegetation. The results of this chapter highlight the significant of earth observation technology to quantifying areas at risk of malaria transmission in line with objective 3 of this study. The following chapter seeks to address how remote sensing can aid in mapping the distribution of tree species *Lippia javanica* (lemon bush) used as an alternative to western medicine for repelling mosquitoes in the VDM.

## 5.6. References

- Adams J. (2010). Vegetation–climate interaction. “How plants make the global environment”, 2nd Edn. Springer, New York. ISBN:978-3-642-00880-1.
- Adams M.L., Philpot W.D., Norvell W.A. (1999). Yellowness index: An application of spectral second derivatives to estimate chlorosis of leaves in stressed vegetation. *International Journal of Remote Sensing* 20: 3663 – 3675.s
- Adeola A.M., Botai J.O., Olwoch J.M., Rautenbach H.C.J., Kalumba A.M., Tsela P.L., Adisa M.O., Wasswa N.F., Mmtoni P., Ssentongo A. (2015). Application of geographical information system and remote sensing in malaria research and control in South Africa: a review. *Southern African Journal of Infectious Diseases* 1: 1 – 8.
- Adeola A.M., Botai O.J., Olwoch J.M., Rautenbach C.J., Adisa O.M., Taiwo O.J., Kalumba A.M. (2016). Environmental factors and population at risk of malaria in

Nkomazi municipality, South Africa. *Tropical Medicine and International Health*, 21(5): 675 – 686.

- Adimi F., Soebiyanto R.P., Safi N., Kiang R. (2010). Towards malaria risk prediction in Afghanistan using remote sensing. *Malaria Journal* 9:125.
- Ahmed A. (2014). GIS and remote sensing for malaria risk mapping, Ethiopia. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Technical Commission VIII Symposium, 09 – 12 Dec 2014, Hyderabad, India* 8: 155 – 161.
- Alimi T., Fuller D.O., Qualls W.A., Herrera S.V., Arevalo-Herrera, Quinones M.L., Lacerda M.V.G., Beier J.C. (2015). Predicting potential ranges of primary malaria vectors and malaria in northern South America based on projected changes in climate, land cover and human population. *Parasites and Vectors* 8: 1 – 16.
- Araujo S. L., Santos J.R., Shimabukuro Y.E. (2000). Relationship between SAVI and biomass data of forest and savannah contact zone in the Brazilian Amazonia. *International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B7*. Amsterdam.
- Baeza A., Bouma M.J., Bobson A.P., Dhiman R., Srivastava H.C., Mascual M. (2011). Climate forcing desert malaria: the effect of irrigation. *Malaria Journal* 10 (90):1 – 10.
- Baldwin R.A. (2009). Use of maximum entropy modelling in wildlife research. *Entropy* 11:854–866.
- Barbet-Massin M., Jiguet F., Albert C.H., Thuiller W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2): 327 – 338.
- Bennie J., Hill M.O., Baxter R., Huntley B. (2006). Influence of slope and aspect on long-term vegetation change in British chalk grasslands. *Journal of Ecology* 94: 355 – 368.
- Berkson J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association* 33: 526 – 542.
- Bernstein L.S., Adler-Golden S., Jin X., Gregor B. (2012). Quick atmospheric correction (QUAC) code for VNIR-SWIR spectral imagery: Algorithm details. *The 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Shanghai*: 1 – 4.
- Bowman W.D. (1989). The relationship between leaf water status, gas exchange and spectral reflectance in cotton leaves. *Remote Sensing of Environment* 30(3): 249 – 255.



- Busby R.(1991). BIOCLIM: A bioclimate analysis and prediction system. *Plant Protection Quarterly* 6.
- Calcagno V., de Mazancourt C. (2010). glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. *Journal of Statistical Software* 34: 1 – 29.
- Ceccato P., Flasse S., Gregoire J. (2002). Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications. *Remote Sensing of Environment* 82: 198 – 207.
- Ceccato P., Flasse S., Tarantola S., Jacquemoud S., Gregoire J. (2001). Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment* 77: 22 – 33.
- Chefaoui R.M., Lobo J.M. (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modeling* 210: 478 – 486.
- Clennon J.A., Kamanga A., Musapa M., Shiff C., Glass G.E. (2010). Identifying malaria vector breeding habitats with remote sensing data and terrain-based landscape indices in Zambia. *International Journal of Health Geographics* 9 (58): 1 – 13.
- Collett D. (1991). Modeling binary data. Chapman and Hall, London.
- Craig M.H., Snow R.W., Le Sueur D. (1999). A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today* 15(3): 105 – 111.
- Dambach P., Machault V., Lacaux J., Vignolles C., Sauerborn R. (2012). Utilization of combined remote sensing techniques to detect environmental variables influencing malaria vector densities in rural West Africa. *International Journal of Health Geographics* 11(8): 1 – 12.
- De Oliveira E.C., do Santos E.S., Zeilhofer P., Souza-Santos R., Atanaka-Santos M. (2013). Geographic information systems and logistic regression for high-resolution malaria risk mapping in a rural settlement of the southern Brazilian Amazon. *Malaria Journal* 12:420.
- ENVI (2009). Atmospheric correction module: QUAC and FLAASH user's guide. Version 4.7. 20AC47DOC.
- Exelis Visual Information Solutions. (2016). Environment for Visualizing Images. Boulder, CO: Exelis Visual Information Solutions.
- Gao B. (1996). NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58: 257 – 266.





- Gerritsen A.A.M., Kruger P., Schim van der Loeff M.F., Grobusch M.P. (2008). Malaria incidence in Limpopo Province, South Africa, 1998 – 2007. *Malaria Journal* 7:1 – 8.
- Gitelson A.A. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology* 160: 271 – 282.
- Gomez-Elipe A., Otero A., van Herp M., Aguirre-Jaime A. (2007). Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997-2003. *Malaria Journal*, 6:129.
- Huete A.R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25(3):295–309.
- Hunt J.R., Rock B.N. (1989). Detection of changes in leaf water content using near-infrared and middle-infrared reflectances. *Remote Sensing of Environment* 30: 43 – 54.
- Ikeda T., Behera S.K., Morioka Y., Minakawa N., Hashizume M., Tsuzuki A., Maharaj R., Kruger P. (2017). Seasonally lagged effects of climatic factors on malaria incidence in South Africa. *Scientific Reports* 7(2458): DOI:10.1038/s41598-017-02680-6.
- Jackson R.D, Slater P.N, Pinter P.J. (1983). Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres. *Remote Sensing of Environment* 13:187–208.
- Jacob B.G., Muturi E.J., Mwangangi J.M., Funes J., Caamano E.X., Muriu S., Shillilu J., Githure J., Novak R.J. (2007). Remote and field level quantification of vegetation covariates for malaria mapping in three rice agro-village complexes in Central Kenya. *International Journal of Health Geographics*, 6(21): DOI: 10.1186/1476-072X-6-21.
- Kabanda, T. and Munyati, C. (2010). Anthropogenic-induced climate change and the resulting tendency to land conflict; The case of the Soutpansberg region, South Africa; Climate Change and Natural Resources Conflicts in Africa, Eds. Donald Anthony Mwiturubani and Jo-Ansie van Wyk. *Monograph No 170: Institute for Security Studies*.
- Kamau L., Munyekenye G.O., Vulule J.M., Lehmann T. (2006). Evaluating genetic differentiation of *Anopheles arabiensis* in relation to larval habitats in Kenya. *Infections, Genetics and Evolution* 7: 293 – 297.
- Kazembe L.N., Kleinschmidt I., Holtz T.H., Sharp B.L. (2006). Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. *International Journal of Health Geographics* 5(41): 1 – 9.

- Kleinschmidt I., Bagayoko M., Clarke G.P.Y., Graig M., Le Sueur D. (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology* 29: 355 – 361.
- Komen K. (2017). Could Malaria Control Programmes be timed to coincide with onset of rainfall? *EcoHealth* 14(2): 259 – 271.
- Komen K., Olwoch J., Rautenbach H., Botai J., Adebayo A. (2015). Long-run relative importance of temperature as the main driver to malaria transmission in Limpopo Province, South Africa: a simple econometric approach. *EcoHealth* 12(1): 131 – 143.
- Lopatin, J., F. E. Fassnacht, T. Kattenborn, and S. Schmidlein. 2017. Mapping plant species in mixed grassland communities using close range imaging spectroscopy. *Remote Sensing of Environment*, 201:12–23.
- Mabaso M.L.H., Vounatsou P., Midzi S., Da Silva J., Smith T. (2006). Spatio-temporal analysis of the role of climate in inter-annual variation of malaria incidence in Zimbabwe. *International Journal of Health Geographics* 5(20): 1 – 9.
- Machault V., Vignolles C., Borchi F., Vounatsou P., Pages F., Briolant S., Icaux J., Rogier C. (2011). The use of remotely sensed environmental data in the study of malaria. *Geospatial Health* 5: 151 – 168.
- Machault V., Vignolles C., Pagès F., Gadiaga L., Gaye A., Sokhna C., Trape J., Icaux J., Rogier C. (2010). Spatial heterogeneity and temporal evolution of malaria transmission risk in Dakar, Senegal, according to remotely sensed environmental data. *Malaria Journal* 9(252): 1 – 14.
- Malahlela O, Cho M.A., Mutanga O. (2014). Mapping canopy gaps in an indigenous subtropical coastal forest using high resolution WorldView-2 data. *International Journal of Remote Sensing* 35: 6397–6417.
- Mutanga O., Skidmore A.K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19): 3999 – 4014
- Moyes C.L., Shearer F.M., Huang Z., Wiebe A., Gibson H.S., Mohd-Azlan J., Brodie J.F., Malaivithnond S., Linkie M., Samejima H., O'Brien T.G., Trainor C.R., Hamada Y., Giordano A.J., Kinnaird M.F., Elyazar I.R.F., Sinka M.E., Vythilingam I., Bangs M.J., Pigott D.M., Weiss D.J., Golding N., Hay S.I. (2016). Predicting the geographical distributions of the macaque hosts and mosquito vectors of *Plasmodium knowlesi* malaria in forested and non-forested areas. *Parasites and Vectors* 9(242): 1 – 12.





- Mpandeli S. (2014). Managing Climate Risks Using Seasonal Climate Forecast Information in Vhembe District in Limpopo Province, South Africa. *Journal of Sustainable Development* 7 (5): 68 – 81.
- National Department of Health. (2011). *Malaria Elimination Strategy 2011 – 2018 in South Africa*. South Africa: National Department of Health (NDoH).
- National Institute for Communicable Diseases (NICD)-National Health Laboratory Services (NHLS) (2017). Malaria in South Africa: An Update. 16(5):1
- Nmor J.C., Sunahara T., Goto K., Futami K., Sonye G., Akweywa P., Dida G., Minakawa N. (2013). Topographic models for predicting malaria vector breeding habitats: potential tools for vector control managers. *Parasites and Vectors* 6: 1 – 13.
- Omumbo J.A., Hay S.I., Goetz S.J., Snow R.W., Rogers D.J. (2002). Updating Historical Maps of Malaria Transmission Intensity in East Africa Using Remote Sensing. *Photogrammetric Engineering & Remote Sensing* 68: 161 – 166.
- Paaijmans K.P., Thomas M.B. (2011). The influence of mosquito resting behaviour and associated microclimate for malaria risk. *Malaria Journal*, 10:183. <https://doi.org/10.1186/1475-2875-10-183>.
- Phillips S.J., Anderson R.P., Schapire R.E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling* 190: 231–259.
- Phillips S.J., Dudík M., Elith J., Graham C.H., Lehmann A., Leathwick J., Ferrier S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19:181 – 197.
- QGIS Development Team. (2016). QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.
- R Development Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Raman J., Morris N., Frean J., Brooke B., Blumberg L., Kruger P., Mabusa A., Raswiswi E., Shandukani B., Misani E., Groepe M., Moonasar D. Reviewing South Africa's malaria elimination strategy (2012-2018): progress, challenges and priorities. *Malaria Journal* 15:438.
- Reason C.J.C., Keibel A. (2004). Tropical cyclone Eline and its unusual penetration and impacts over the southern African mainland. *Weather Forecast* 19(5): 789 – 805.



- Ricotta E.E., Frese S.A., Choobwe C., Louis T.A., Shiff C.J. (2014). Evaluating local vegetation cover as a risk factor for malaria transmission: a new analytical approach using ImageJ. *Malaria Journal*, 13(94): DOI: 10.1186/1475-2875-13-94.
- Ripley B.D. (2003). Selecting amongst large classes of models. Lecture, URL <http://www.stats.ox.ac.uk/~ripley/Nelder80.pdf>
- Sinka M.E., Rubio-Palis Y., Manguin S., Patil A.P., Temperly W.H., Gething P.W., Van Boeckel T., Kabaria C.W., Harbach R.E., Hay S.I. (2010). The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and biologic précis. *Parasites and Vectors* 3: 1 – 26.
- Statistics South Africa. (2016). Population census 2016. Vhembe District Municipality. [http://cs2016.statssa.gov.za/?page\\_id=270](http://cs2016.statssa.gov.za/?page_id=270)
- Stockwell D., Peters D. (1999). The GARP modelling system. Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, 13: 143 – 158.
- Tonnang H.E.Z., Kangalawe R.Y.M., Yanda P.Z. (2010). Predicting and mapping malaria under climate change scenarios: the potential distribution of malaria vectors in Africa. *Malaria Journal* 9(111): 1 – 10.
- Tucker C.J. (1980). Remote sensing of leaf water content in the near infrared. *Remote Sensing of Environment* 10(1): 23 – 32.
- Wayant N.M., Maldonado D., Rojas de Aria A., Cousiño B., Goodin D.G. (2010). Correlation between normalized difference vegetation index and malaria in a subtropical rain forest undergoing rapid anthropogenic alteration. *Geospatial Health* 4: 179 – 190.
- World Health Organization (WHO) (2016). World malaria report 2016. Geneva: World Licence:CC BY-NC-SA 3.0 IGO.
- Yang X., Skidmore A.K., Melick D.R., Zhou Z., Xu J. (2006). Mapping non-wood forest product (matsutake mushrooms) using logistic regression and a GIS expert system. *Ecological Modelling*, 198: 208 – 218.



## CHAPTER 6<sup>5</sup>:

# Mapping the spatial distribution of *Lippia javanica* (Burm. f.) Spreng using Sentinel-2 and SRTM-derived topographic data in malaria endemic environment

Oupa E. Malahlela <sup>a,b\*</sup>, Adjorlolo C <sup>c.</sup>, Olwoch J.M.<sup>a,d</sup>

<sup>a</sup> Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

<sup>b</sup> South African National Space Agency (SANSA), Earth Observation Directorate, Pretoria 0001, South Africa

<sup>c</sup> New Partnership for Africa's Development (NEPAD) Agency, 230 15<sup>th</sup> Road, Midrand, South Africa

<sup>d</sup> Southern African Science Service Center for Climate Change and Adaptive Land Management (SASSCAL), Windhoek 91100, Namibia

***This paper relates to objectives 4 of the thesis.***

---

<sup>5</sup>This chapter is based on the manuscript titled "*Mapping the spatial distribution of Lippia javanica (Burm. f.) Spreng using Sentinel-2 and SRTM-derived topographic data in malaria endemic environment*" **Ecological Modelling**, 392: 147-158.

## Abstract

*Lippia javanica* (*L. javanica*) is one of the commonly used ethnobotanical plant species for controlling malaria globally. Accurate mapping of *L. javanica* species is important for malaria control interventions that require geospatial information for the assessment of malaria distribution and monitoring especially in communities that have limited access to western malaria medicine. Currently, high spatial resolution information pertaining to the distribution and habitat suitability of *L. javanica* species is very rare. The high resolution mapping could assist in identifying potential niche areas of ethnobotanically important species and to facilitate community health and wellness against malaria. In this chapter, we tested the ability of high spatial resolution Sentinel-2 (S2) derived variables and Shuttle Radar Topography Mission (SRTM)-derived topographic variables to predict the distribution of *L. javanica* in the Vhembe District Municipality (South Africa). The relationship between remote sensing variables and the occurrence data of *L. javanica* was assessed using coefficient of determination ( $R^2$ ). We compared three commonly used species distribution models (logistic regression, Maxent and ensemble models) to derive the best possible subsets of environmental predictors, and to produce the species distribution map that could aid in identifying areas where *L. javanica* occurs for use against malaria vectors. The probability threshold of  $> 0.6$  on the predicted data and the area under curve of the receiver operating curve (ROC) were used as additional validation methods, using independent validation dataset. The results showed a superior performance of weighted ensemble model, which yielded higher overall accuracy (86.7%, AUC = 0.89) than both logistic regression (77.7%, AUC = 0.79) and Maxent (80.0%, AUC = 0.83). The indices derived from the Sentinel's red edge bands were the most contributory variables in both logistic regression and Maxent. The normalized difference averaged red edge vegetation index (NDARVI) and the normalized difference red edge1 vegetation index (NDVIre) contributed 39.25% ( $p = 0.0002$ ) and 32.50% in LR and Maxent models respectively. Slope was the most significant SRTM-derived variable correlated to *L. javanica* occurrence in all models. The results of this chapter show that high resolution S2 data can be used to map hardy shrub species at higher accuracies using ensemble model. The derived occurrence map of *L. javanica* could assist in updating the currently coarse resolution distribution map of species required for its aromatic ecological service against malaria vectors.

**Keywords:** *Lippia javanica*, malaria; Sentinel-2; SRTM; Vhembe District Municipality

---

## 6.1 Introduction

In 2015, the World Health organization has developed a global technical strategy for malaria 2016–2030 (GTS) (5), endorsed by the World Health Assembly. This strategy promulgated malaria elimination efforts targeted for 2030, with milestone for measuring progress in both 2020 and 2025 (WHO, 2018). Globally, it was recommended that all countries, particularly in Africa, set their own national or subnational targets for accelerating efforts aimed at eliminating malaria transmission and preventing re-infections. Many countries, especially in Africa have

developed their own malaria control plans and subsequent implementations resulted in the decrease in the number of malaria transmissions (Pierre-Louis *et al.*, 2018). For example, many of these countries have adopted the use of insecticide treated nets (ITNs) and indoor residual sprays (IRS), which have become cornerstones for malaria control programme (WHO, 2008; Lengelar, 2004). In most African communities, the use of repellents of plant origin has been one of the most effective methods for controlling malaria for many years (Kweka *et al.*, 2008), because of their cheaper costs of treatment where such plant species are commonly found, and have shown to contain low toxicity to humans and animals (Ansari *et al.*, 2000). Many plant species have been effective alternative for repelling malaria vectors, and these plant species include *Lantana camara*, *Omicum americanum*, *Azadirachta indica*, and *Lippia javanica* (Seyoum *et al.*, 2002; Mabogo, 1990).

In other parts of Africa and Indian Subcontinent, *Lippia javanica* (Burm. f) Spreng is still one of the commonly used medicinal plants for controlling malaria and for ethnoveterinary purposes (Maroyi, 2017; Lukwa, 1994; Samie *et al.*, 2005; Kumar and Dash, 2012). In Southern Africa, *L. javanica* is one of the four indigenous *Lippia* species (Verbenaceae) that occur as erect woody shrubs of approximately 2 meters in height (Viljoen *et al.*, 2005). This species is one of the widely used ethnobotanical plants for control of malaria in Zimbabwe and South Africa. For example, in the Vhembe District Municipality of South Africa, the Venda people use *L. javanica* leaf infusions as prophylactic against malaria (Mabogo, 1990; Maroyi, 2017). Today, *L. javanica* is still widely used in many communities in Africa as a means to keep both nuisance and malaria transmitting mosquitoes at bay (Mavundza *et al.*, 2011). Because of *L. javanica* significance as an antimalarial species in Southern Africa, accurate knowledge of species occurrence and habitat delineation is fundamental for continued fight against malaria infections in rural communities. This knowledge will further enhance our understanding of suitable habitats for *L. javanica* species amidst changes in climatic conditions, which radically alters species distribution and functions (DeHayes *et al.*, 2000).

Delineating habitats of plant species conventionally relies on ground-based surveys which are the most common means for species detection (Legendre *et al.*, 2002; Gogol-Prokurat, 2010). Through ground-based surveys, efforts are made to locate plant species of interest along pre-defined transects, mostly biased towards readily accessible sampling areas (Malahlela *et al.*, 2015; Edwards *et al.*, 2007). The challenge with ground-based surveys rests in the difficulty to account for plant species that occur in inaccessible areas, due to hostile terrain and logistical constraints associated with ground-based surveys. Moreover, although mapping of species distribution in this manner is mostly accurate, the ground-based surveys, whether of

plant or animal communities, are often time-consuming and costly (Rocchini *et al.*, 2010). In order to overcome the limitations associated with these surveys, highly efficient and robust methods are mostly recommended for species distribution studies.

Remote sensing offers fast and efficient alternative approach for mapping the potential distribution of plant species. This is because remote sensing data is acquired for much larger areas at various spatio-temporal and spectral resolutions (Kempeneers *et al.*, 2011). The spatio-temporal resolution allows for the detection of plant functioning at species (Cho *et al.*, 2015; Kganyago *et al.*, 2018), community (Madonsela *et al.*, 2018), or regional levels (Gould, 2000) over time. On the other hand, the spectral configuration of remote sensing instruments allows for discrimination amongst various plant species and to characterize topographical and environmental features that are associated with habitat of species of interest (Peng *et al.*, 2018; Álvarez-Martínez *et al.*, 2017). For example, Wakie *et al.* (2014) used the time-series 250 m moderate resolution imaging spectroradiometer (MODIS) data to map the current and potential distribution of shrubby *Prosopis juliflora* in the Afar Region of Ethiopia. However, the challenge with coarse resolution habitat mapping lies in the inability to resolve finer details (< 20 m) relating to characteristics of species of interest's suitable habitat. In order to circumvent this problem, some authors have advocated for the use of high spatial resolution remote sensing dataset. For example, Fuller (2005) successfully mapped the distribution of invasive *Melaleuca quinquenervia* species in South Florida using the 4 meter spatial resolution IKONOS imagery. The use of high spatial resolution imagery such as IKONOS has its own inherent limitations. One such limitations is the lack of narrow-band spectral configuration that has shown high sensitivity to subtle variations in vegetation and habitat characteristics (Mutanga and Skidmore, 2004; Malahlela *et al.*, 2015). Elsewhere it was documented that the narrow-band hyperspectral data show greater potential for mapping trees species distribution (He *et al.*, 2011). Therefore, the remote sensing data which offers the advantage of high spatial resolution multispectral and narrow-band hyperspectral band configuration is necessary for mapping the distribution of *L. javanica* species.

The launch of Sentinel-2 (S2) satellite in June 2015 by the European Space Agency (ESA) has opened opportunities for high resolution vegetation habitat characterization. The vegetation indices (VIs) derived from S2 broadband wavelengths (visible-near infrared region) have been used to retrieve vegetation biophysical/chemical parameters (Clevers *et al.*, 2017). In contrast, some studies have shown that the narrow-band VIs outperform their broadband counterparts (de Oliveira *et al.*, 2017; Ramoelo and Cho, 2018; Korhonen *et al.*, 2017), due to ability to compensate for signal saturation problem (Frampton *et al.*, 2013). Although the



broadband vegetation indices such as the normalized difference vegetation index (NDVI) have been used for species mapping (West *et al.*, 2016a; Pau *et al.*, 2012), the use of narrow-band indices such as those derived from red edge region have rarely been used for species distribution mapping. Additionally, the topographical variables such as digital elevation model (DEM), aspect and slope have been widely used to describe topography associated with the species' suitable habitat in species distribution mapping (Guisan and Zimmermann, 2000; Zeng *et al.*, 2017). However, the successful mapping of *L. javanica* requires the model with high sensitivity to the potential habitat of the species under consideration.

Most of the currently available species distribution models (SDMs) offer a versatile approach to empirically relate the satellite-derived environmental variables with distribution of plant species (He *et al.*, 2015; Dudov, 2017). Some of the most commonly used SDMs include the logistic regression (Padalia *et al.*, 2010; Lemke and Brown, 2012) and the maximum entropy model (Maxent) (Chahouki and Sahragard, 2016; Dudov, 2017; Ullerud *et al.*, 2016). By comparison, the performance of Maxent model is comparable or superior to other SDMs such as the logistic regression model (Elith *et al.*, 2011). Although these models are popular and effective for predicting areas of suitable habitat in unsampled locations, they have their limitations. For example, despite its success in species distribution modelling, Maxent cannot be used to model a species fundamental niche, which concerns a full range of environmental conditions a species might occupy, not just environmental conditions where species is most likely to be found (Drake, 2014; Pulliam, 2000). On the other hand, the inability of logistic regression model to account for imperfect detection which often leads to biased estimation of habitat relationships is well-documented (Martin *et al.*, 2005). These limitations associated with each SDM can be circumvented by ensemble SDM, that combines the strength of both logistic regression and Maxent thus minimizing errors associated with each model (Araujo and New, 2007). Ensemble model is a combination of individual SDM with differing structure, explanatory variables, and the data sources for the purpose of describing species' relationship with the environment.

The aim of the current chapter was to map the potential distribution of *L. javanica* (lemon bush) shrub species in a malaria-endemic area, using high resolution satellite remote sensing. The objectives of this chapter were to (i) assess the contribution of broad-band, narrow-band vegetation indices and topographical variables in logistic regression and Maxent model, (ii) compare the performance accuracies of logistic regression, Maxent and ensemble model for predicting *L. javanica* distribution, and (iii) spatially map the distribution of *L. javanica* by using a more robust approach.

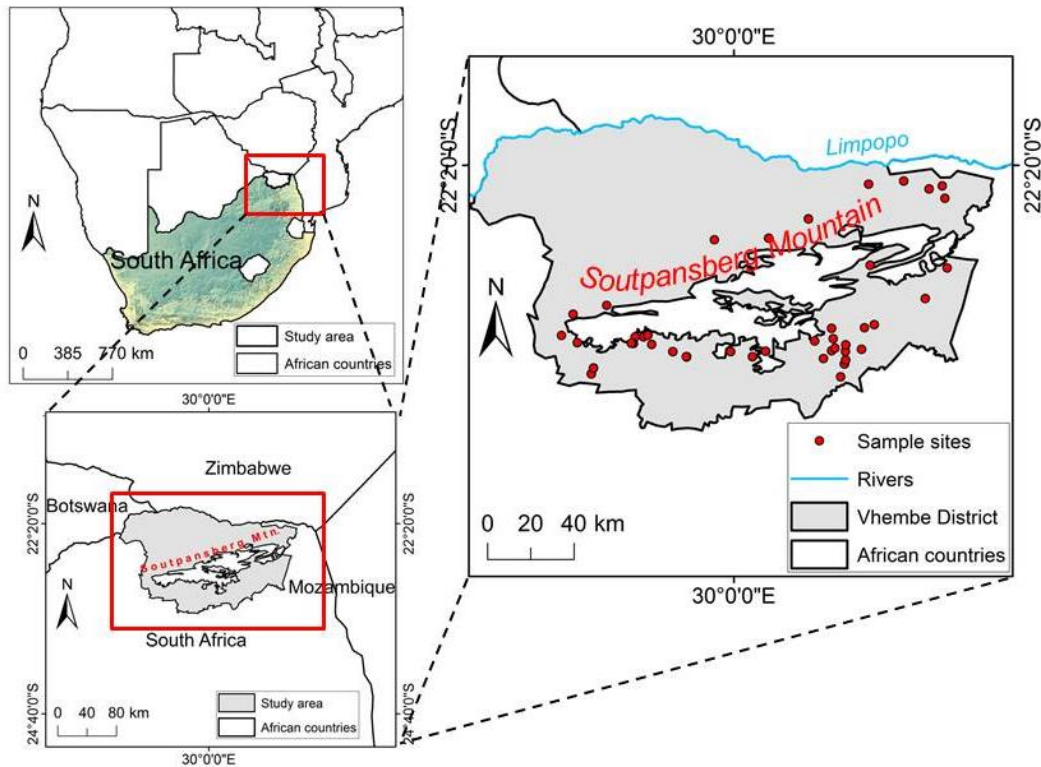




## 6.2 Materials and methods

### 6.2.1 Study area

The study was conducted at the Vhembe District Municipality (VDM), which is the northernmost district of South Africa (Figure 6.1). The study area is located at the geographic coordinates of 23°40'S and 30°00'E (Malahlela *et al.*, 2018). The VDM comprises of four local municipalities, (i.e. Thulamela LM, Musina LM, Makhado LM and Collins Chabane LM). The total human population in the VDM is currently over 1.3 million (Stats SA, 2012; Malahlela *et al.*, 2018), with predominantly Venda people occupying the western, and north and north-eastern part of the VDM. The study area forms administrative border between South Africa and Zimbabwe, with the Limpopo River being the most notable natural boundary between these two countries. It comprises of varying topography, with diverse floral and faunal biodiversity, with *Adansonia digitata* (baobab tree) as the well-known tree species occurring in this region. The VDM is situated in the tropical savannah biome, which comprise of mixture of C<sub>4</sub> grasses and closed-canopy trees (Lehmann *et al.*, 2011; Asner *et al.*, 2004). The vegetation density decreases from east to west, with high vegetation density found in the eastern part of the study area, mainly in response to rainfall patterns in South Africa (Bond *et al.*, 2002). The study area receives annual rainfall of 820 mm (Mpandeli, 2014), and varied localized rainfall frequencies and quantities occurring closer to Soutpansberg Mountain (Kabanda and Munyati, 2010). It is for this reason that Matiwa in the VDM had recorded the average rainfall of 2004 mm (over 60 year calculation), which is the highest ever recorded rainfall in South Africa (SAWS, 2018).



**Figure 6.1.** Limits of the study area (Vhembe District Municipality) and the sample locations used for modelling (shown by red dots).

Due to semi-aridity and tropical conditions prevailing in different parts of the study area, malaria transmission exhibit seasonal variation, with high malaria incidences usually occurring between September and April (Adeola *et al.*, 2017). Most communities in the study area use *L. javanica* species for controlling mosquito bites (Phaswane and Masevhe, 2018), thus preventing potentially large scale malaria transmissions.

### 6.2.2 Field data

Field data collection was undertaken in August 2017 for 21 days, covering most parts of the VDM. The field data was collected purposively across four vegetation cover classes, (i) shrubland, (ii) grassland, (iii) cropland and (iv) woodland. Data collection was done within a 10 m × 10 m sub-plot in a larger 30 m × 30 m plot, where species sightings were recorded as either present or absent. The *L. javanica* was spotted mostly under semi-open canopy tree species such as young *Acacia tortilis*, and usually occurred as single stand of shrub. The geo-locations of each plot and species sightings were recorded using a handheld Garmin eTrex 20™ global positioning system (GPS) with the maximum positional accuracy of 3m. A total number of 151 ( $n=151$ ) samples were collected for analysis.

### **6.2.3 Remote sensing/topographical data**

S2 data was used for the study. The imagery used were acquired between August 2017 and September 2017, and were downloaded free of charge from Copernicus Science hub (<https://scihub.copernicus.eu/dhus/#/home>). The S2 was first launched in June 2015 by the European Space Agency. It has spectral bands ranging from coastal aerosol (443 nm) to the shortwave infrared (2190 nm). The blue (490 nm), green (560 nm), red (660 nm) and near infrared (842 nm) are available at 10 m spatial resolution; the red edge1 (705 nm), red edge2 (740 nm), red edge3 (783 nm), red edge4 (865 nm), shortwave1 (1610 nm), and shortwave2 (2190 nm) are available at 20 m spatial resolution, while the coastal (443 nm), water vapour (945 nm) and cirrus band (1375 nm) are available at 60 m spatial resolution. The elevation data, slope and aspect were derived from the Shuttle Radar Topography Mission (SRTM) with 30 m spatial resolution.

### **6.2.4 Data pre-processing**

Data pre-processing was done in three sets, (i) the S2 data pre-processing, (ii) the SRTM data pre-processing, and (iii) the field data pre-processing.

#### **6.2.4.1 Sentinel-2 data**

Pre-processing of S2 data involved spatial resampling of wavelengths to address the objectives of the study. Firstly, a total of ten spectral bands ( $n = 10$ ) were selected based on their relationship with *L. javanica* distribution. These bands are centered at 490 nm (absorbed by chlorophyll), 560 nm (sensitive to plant health such as greenness), 660 nm (absorbed by chlorophyll), 705 nm – 865 nm (sensitive to subtle variations of vegetation chlorophyll), 842 nm (sensitive to leaf mass and chlorophyll content), 1610 nm (sensitive to vegetation moisture content) and 2190 (sensitive to vegetation moisture and soil minerals) (Malahlela *et al.*, 2014; Ustin *et al.*, 2009). Secondly, the 10-band S2 data was subjected to atmospheric correction using Sen2Cor code in order to minimize atmospheric effect on the target spectra. Sen2cor uses a large database of look-up tables (LUT) derived using an atmospheric radiative transfer model based on libRadtran1 (Müller-Wilm, 2016). Thirdly, the atmospherically corrected bands were stacked in Quantum GIS (QGIS Development Team, 2018) software, resulting in six single multispectral imagery with 10 bands each and 20 m spatial resolution. The subsequent processing involved spatial resampling of images to 10 m spatial resolution, and this was accomplished in Sentinels Applications Platform software (SNAP), using nearest neighbour resampling method. Lastly, the 10 m imagery were mosaicked and clipped to cover the entire study area.

The broad-band VIs representing photosynthetic pigments (NDVI), senescent vegetation and soil (soil-adjusted vegetation index, optimized soil-adjusted vegetation index), vegetation and landscape water content (normalized difference water index) and herbaceous biomass (difference vegetation index, simple ratios) were computed (Hill, 2013). These indices were selected on the basis of their individual sensitivity to vegetation characteristics and their ability to minimize the soil brightness. For example, the DVI, NDVI and GRVI have shown to be sensitive to vegetation green-up, biomass and biophysical characteristics such as diversity (Lecain *et al.*, 2000; Madonsela *et al.*, 2018). However due to the vegetation composition nature of our study area (very dense to very sparse vegetation with exposed soil background), both SAVI and OSAVI were selected to compensate for the effect of high bare soil-to-vegetation cover ratio. The intention was to derive minimum number of indices which takes into account the landscape dynamics in terms of vegetation biochemical and biophysical properties of *L. javanica* habitats at the study area. Moreover, two narrow-band red edge indices (normalized difference normalized red edge vegetation index, Normalized difference averaged red edge vegetation index) were computed and tested for the study because of their ability to detect subtle variations in vegetation characteristics (Table 6.1).

**Table 6.1:**

Broad-band and narrow-band spectral indices used in modelling the distribution of *L. javanica* at the study area.

No.	Spectral index	Equation	References
1	Difference vegetation index (DVI)	$DVI = R_{842} - R_{660}$	Tucker (1979)
2	Normalized difference vegetation index (NDVI)	$NDVI = \frac{(R_{842} - R_{660})}{(R_{842} + R_{660})}$	Rouse <i>et al.</i> (1974)
3	Normalized difference red edge1 vegetation index (NDVire)	$NDVire = \frac{(R_{842} - R_{705})}{(R_{842} + R_{705})}$	Gitelson and Merzlyak (1994)
4	Normalized difference averaged red edge vegetation index (NDARVI)	$NDARVI = \frac{(R_{842} - a)}{(R_{842} + a)}$	In this chapter
5	Soil-adjusted vegetation index (SAVI)	$SAVI = \left( \frac{R_{842} - R_{660}}{R_{842} + R_{660} + L} \right) \cdot (1 + L)$	Huete (1988)



6	Optimized soil-adjusted vegetation index (OSAVI)	$\text{OSAVI} = \left( \frac{R_{842} - R_{660}}{R_{842} + R_{660} + 0.16} \right)$	Rondeaux <i>et al.</i> (1996)
7	Green ratio vegetation index (GRVI)	$\text{GRVI} = R_{560} - R_{660}$	Gamon and Surfus (1999)
8	Normalized difference water index (NDWI)	$\text{NDWI} = \frac{(R_{842} - R_{1610})}{(R_{842} + R_{1610})}$	Gao (1996)
9	Simple ratio1 (SR1)	$\text{SR1} = \frac{R_{842}}{R_{490}}$	Gitelson and Merzlyak (1997)
10	Simple ratio2 (SR2)	$\text{SR2} = \frac{R_{842}}{R_{660}}$	Gitelson and Merzlyak (1997)

---

$L$  is soil-line coefficient of 0.5;  $a$  is the mean reflectance of four red edge wavelengths.

---

#### 6.2.4.2 SRTM data

The SRTM digital elevation (DEM) data was downloaded from the United States Geological Surveys' Earth Explorer (<https://earthexplorer.usgs.gov/>). The DEM was clipped to the VDM boundary extent. The DEM was resampled to a 10 m spatial resolution in QGIS, for use in Maxent and logistic regression models. This was done by assigning the same number of spatial dimensions for both longitude (X) and latitude (Y) so as to match those of a 10 m S2 multispectral image. It was from the same DEM where additional environmental datasets such as slope and aspect were derived.

#### 6.2.4.3 *Lippia javanica* field data

The preparation of field point data was conducted in Microsoft Spreadsheet where species locations, occurrence, and corresponding environmental data were appended. A total of 151 presence-absence sampling points ( $n = 151$ ) was collected during field survey. The field data was later imported into ArcMap vers.10 (ESRI Inc, Redlands) as a shapefile for visualizing preliminary spatial distribution of species occurrence. All other environmental/remote sensing data were appended on the field data which formed part of the standard dataset to be used for *L. javanica* species mapping.

#### 6.2.5 Modelling strategy

The correlation between satellite derived variables and the presence/absence of *L. javanica* was assessed through the use of Pearson correlation coefficient ( $r$ ). For predicting the distribution of *L. javanica* species, the logistic regression (Hosmer *et al.*, 2013) and Maxent (using Maxent version 3.3.3; Phillips *et al.*, 2006) were tested in the current study. A standard

field data was with presence-absence data was used in logistic regression while only the presence records and background values were used for model fitting in Maxent. It was randomly split into 70% ( $n = 106$ ) and 30% ( $n = 45$ ) for model calibration and independent model validation respectively. Logistic regression was used to link the occurrence of *L. javanica* with the set of environmental variables relating to vegetation greenness, chlorophyll, canopy moisture, vegetation biomass, aspect, elevation and slope. Logistic regression is a form of generalized linear models (GLM) which relates the binary response outcome (presence-absence) to a linear combination of numerical and categorical variables (Hosmer and Lemeshow, 2000). It is given by equation (6.1) as:

$$y_i = \frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{j=1}^k \beta_n x_{nj}\right)\right]} \quad (6.1)$$

where  $y_i$  is the probability of *L. javanica* occurrence (1 or 0),  $x_i$  is the environmental variable at the  $j^{\text{th}}$  location,  $\beta_n$  is the coefficient of  $x_n$ ,  $\beta_0$  is an intercept, and exp is the exponential function of the regression. An automated procedure for selecting variables and final model was adopted because it reduces computation time and tedious modelling (Malahlela *et al.*, 2018). A stepwise logistic regression approach was adopted through glm2 and MASS packages in R software to minimize multi-collinearity and model over-fitting (Collet 1991). The final model was selected on the basis of lowest Aikake's Information Criterion (AIC) and variable significance. The significant model was ultimately used for spatial mapping of *L. javanica* species at the study area. On the other hand, the Maxent algorithm (Phillips *et al.*, 2006) was also tested for predicting the probability of occurrence of *L. javanica* species. The modelling was done within the constraints of the listed environmental and the S2 datasets. Maxent automatically includes variables interactions and can consider continuous and categorical predictor variables (West *et al.*, 2016b). The samples with data (SWD) file format was used (a comma delimited) as input for presence localities of *L. javanica* species. The same number of presence points used in the logistic regression model was used in Maxent, with the exception that the absence points were replaced with the background values equivalent to the removed absence points. Maxent algorithm was implemented with the default regularization so as to avoid too complex a model. For the ensemble model both Maxent and logistic regression models were combined to form one new model, based on simple weighted averaging (mean). The weighted mean of the resultant combined model is derived from equation (6.2) as:

$$(y)_{ens} = \frac{\sum_{i=1}^n (x_i w_i)}{\sum_{i=1}^n w_i} \quad (6.2)$$



where  $(y)_{ens}$  is the ensemble model,  $w$  is the allocated weights of values of  $x$  in models (1 and 2). The average weights of 0.55 and 0.45 were assigned to individually high and low performance models respectively, such that the ensemble model is biased towards model that performs better.

### 6.2.6 Model evaluation

The independent field dataset ( $n = 45$ ) was used to assess the accuracy of each predictive model. Although the area under curve (AUC) provide a useful measure of a model to discriminate between species presence and absence, additional metrics such as threshold selection can be used based on the study objectives (West *et al.*, 2016b). In the current study the predicted probability of *L. javanica* occurrence ( $y$ ) through each model ranged from 0 to 1, with increasing likelihood of species occurrence towards a value of 1. Therefore, in the current study the threshold of 0.6 was used as an optimum cut-off value, since a value of 0.5 usually represents prediction by random chance (Baldwin, 2009). A  $2 \times 2$  contingency tables (with rows indicating predicted cases and columns indicating measured cases) were drawn for the threshold value of  $> 0.6$  for each model. The overall accuracy, which is the proportion of the correctly classified cases over the total number of cases in the validation dataset, was also used as a means to validate each model (Malahlela *et al.*, 2015). The receiver operating curves (ROC) were also used to assess the robustness of a binary classifier.

## 6.3 Results

The preliminary results of descriptive statistics table (Table 6.2) show that *L. javanica* tends to be present in areas of higher elevation (altitude), mainly in the south facing slopes, and at locations where vegetation exhibited increasing vigour and biomass. Regarding vegetation indices, the DVI showed high within class dispersion from the means of both *L. javanica* presence and absence sites (16.2; 19.2), indicating how this variable may impact of species prediction in SDM's. The correlation analysis has shown that S2's NDARVI and the slope derived from SRTM were positively and significantly correlated to *L. javanica* occurrence ( $r = 0.48, p=0.01$ ;  $r = 0.47, p = 0.01$  respectively) (Table 6.3). On the other hand, the NDVI<sub>ire</sub> and elevation were the second most significant variables correlated to the species distribution ( $r = 0.29, p = 0.05$ ;  $r = 0.27, p = 0.01$  respectively). In all the selected predictor variables, the GRVI was found to be the least and insignificant correlated variables to the *L. javanica* presence/absence ( $r = -0.04, p = 0.1$ ). The NDWI exhibited the significant and negative correlation to the species distribution ( $r = -0.26, p = 0.01$ ).



**Table 6.2:**

Descriptive statistics of S2 and SRTM-derived variables at presence and absence sites for *Lippia javanica* species.

Variables measured	<i>Lippia javanica</i> presence sites			<i>Lippia javanica</i> absence sites		
	Mean	Std. dev	Range	Mean	Std.dev	Range
Elevation (m)	810.8	182.1	545.0 – 1063.0	694.6	186.9	333.0 – 964.0
Slope (rad.)	71.2	28.9	27.0 – 114.7	38.6	26.4	6.4 – 117.4
Aspect (°)	184.8	87.9	14.0 – 328.0	165.2	100.9	0.0 – 350.5
NDVI	0.53	0.20	0.22 – 0.88	0.48	0.25	0.16 – 0.93
NDVIre	0.44	0.09	0.06 – 0.23	0.33	0.19	0.11 – 0.74
DVI	26.7	16.2	9.8 – 58.6	22.5	19.2	2.9 – 77.2
SAVI	0.26	0.10	0.11 – 0.43	0.24	0.13	0.08 – 0.46
OSAVI	0.53	0.20	0.22 – 0.86	0.48	0.25	0.16 – 0.93
GRVI	- 0.04	0.38	- 0.50 – 0.51	- 0.01	0.24	- 0.50 – 0.46
NDARVI	0.15	0.06	0.06 – 0.23	0.09	0.03	0.02 – 0.17
NDWI	- 0.14	0.35	- 0.54 – 0.24	0.05	0.27	- 0.64 – 0.58
SR1	30.4	13.9	15.4 – 58.9	26.8	18.0	5.32 – 81.8
SR2	26.1	10.7	15.6 – 52.8	22.5	14.1	5.13 – 55.1

**Table 6.3:**

Correlations between satellite data and the *L. javanica* presence/absence data ( $n = 151$ )

	Variable	Correlation	Significance
SRTM data	Slope	0.47	**
	Elevation	0.27	**
	Aspect	0.10	NS
Sentinel-2 data	NDARVI	0.48	**
	NDVIre	0.29	*
	SR2	0.12	NS
	DVI	0.10	NS
	NDVI	0.10	*
	SR1	0.09	NS
	SAVI	0.09	*



OSAVI	0.09	NS
GRVI	-0.04	NS
NDWI	-0.26	**

Significance (*p*) codes: (\*) = 0.05; (\*\*) = 0.01; NS = not significant

The results of a stepwise logistic regression are shown in table 6.4. A total of six (6) variable (mostly significant) formed part of the final logistic regression model. The NDARVI, Slope, and Elevation were significantly and positively correlated to *L. javanica* occurrence ( $p < 0.001$ ). On the other hand, the DVI and Aspect were also positively correlated to *L. javanica* occurrence, while NDVI exhibited negative but significant correlation ( $p < 0.05$ ). Figure 6.2 shows summary of variable contribution into final model, with NDARVI being the highest contributory variable (39.25%). Furthermore, using the independent validation dataset, the logistic regression yielded overall classification accuracy of 77.7% at a pre-defined threshold of  $> 0.6$ . On the other hand, Maxent model has resulted in overall classification accuracy of 80.0%. In Maxent, the highest contributory variables were the NDVIre (32.50%), Slope (28.6%) and Elevation (24.10%) respectively (Figure 6.2). The combined ('ensemble') model yielded the overall classification accuracy of 86.7%, which is approximately 9% improvement to the prediction made using logistic regression (77.7%), while Maxent was the second best predictive model in the current study. A  $2 \times 2$  error matrix table for all three models is shown in table 6.5. The ROC curves of all models are shown in figure 6.3. Generally, all models were fairly able to predict the occurrence of *L. javanica* species, with the area under ROC curve  $> 0.7$  (Figure 6.3).

**Table 6.4:**

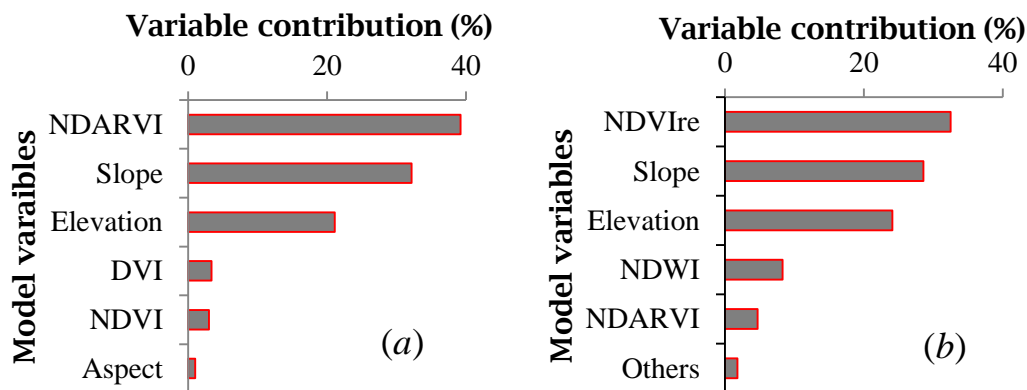
The final predictive model selected using stepwise binary logistic regression

Variable	Estimate	Std. error	P-value
(Intercept)	- 19.599	4.531	0.0001***
NDARVI	49.909	13.492	0.0002***
Slope	0.046	0.014	0.0008***
Elevation	0.014	0.041	0.0003***
DVI	0.136	0.055	0.0131*
NDVI	- 9.399	4.497	0.0366*
Aspect	0.001	0.004	0.7296

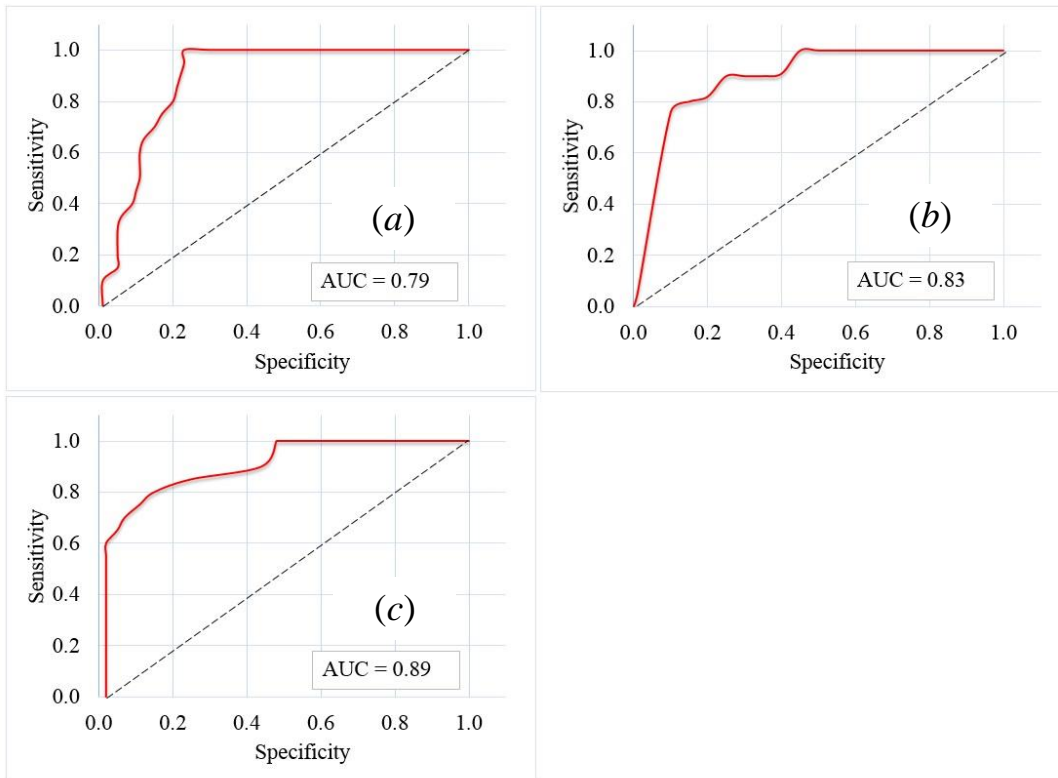
Significance ( $p$ ) codes: (\*) = 0.05; (\*\*) = 0.01; (\*\*\*) = 0.001. AIC = 62.77

### 6.3.1 Predictive maps

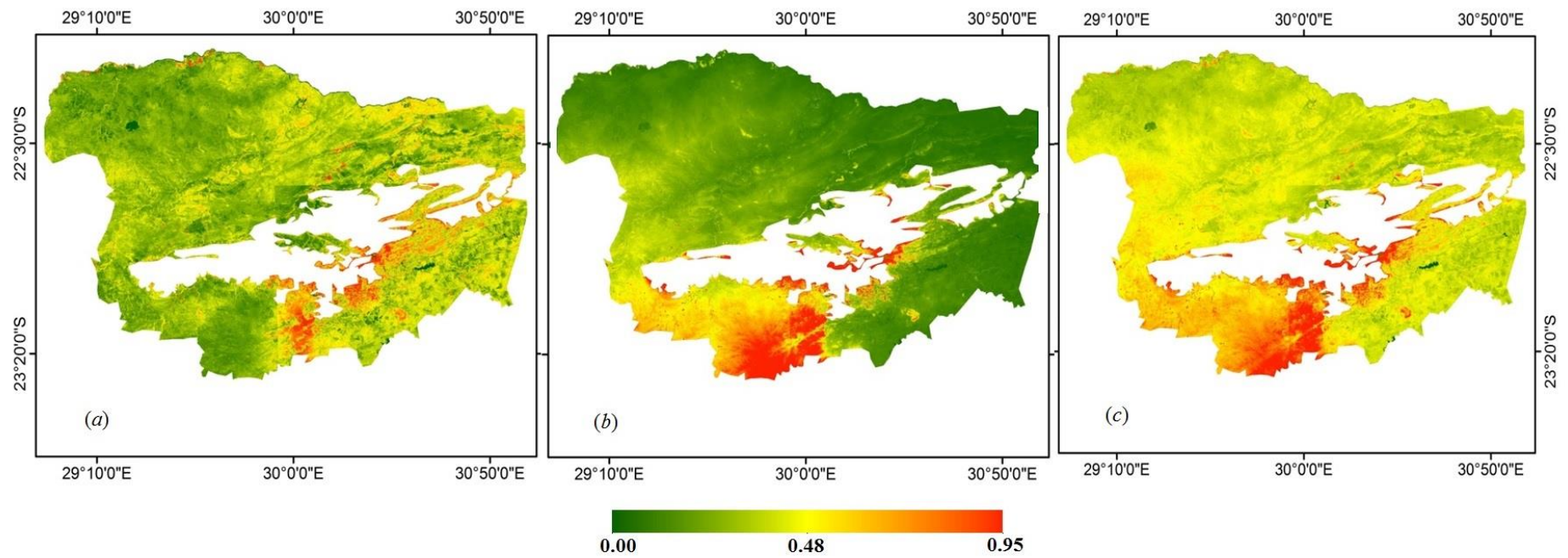
The predicted maps of the study area through Maxent, logistic regression and ensemble had similar spatial pattern (Figure 6.4) especially for presence probabilities of *L. javanica*. High likelihood of *L. javanica* occurrence is predicted in the south-central part of the study area by all 3 models. The minimum predicted value (absence) was 0.0 while 0.95 was the maximum predicted occurrence value in each model. In Maxent, the predicted presence distribution of *L. javanica* exhibited a narrow range, than in logistic regression.



**Figure 6.2:** Analysis of variable importance for both logistic regression (a) and Maxent (b) predictive models ( $n = 106$ ).



**Figure 6.3.** ROC curves of different models used in this chapter (a) logistic regression, (b) Maxent and (c) ensemble model. The dotted line depicts a line of no-discrimination (random guess) while the red line indicates sensitivity and specificity at various threshold levels.



**Figure 6.4.** Comparison of predictions for *L. javanica* using (a) logistic regression model, (b) Maxent and (c) the weighted ensemble modelling. The green colour indicates areas of low probability of occurrence while the red colour corresponds to areas of high species occurrence probability.

**Table 6.5:**

Predicted occurrence vs. observed occurrence of *L. javanica* ( $n = 45$ ) using three SDM's at probability threshold  $> 0.6$ .

		presence	absence	total
<b>Logistic</b>	presence	1	2	3
	absence	8	34	42
<b>Maxent</b>		presence	absence	
	presence	1	0	1
	absence	9	35	44
<b>Ensemble</b>		presence	absence	
	presence	3	0	3
	absence	6	36	42

## 6.1 Discussion

In this chapter, the environmental distribution of *L. javanica* was mapped through the use of Sentinel-2 data. In ecology, the control of malaria by targeting the elimination of mosquitoes is not necessary because at any time once the vector that incubates *P. falciparum* has been eliminated, a more potent one emerges (Killen *et al.*, 2013). This is particularly exacerbated by climate change which increases the risk of malaria re-emergence in endemic and epidemic areas (Ivanescu *et al.*, 2016). The re-emergence of malaria has been reported elsewhere (Sharma, 1996) due to shortage of the dichlorodiphenyltrichloroethane (DDT) commonly used for malaria vector control. Although the DDTs play a crucial role for rapid control against malaria in many countries, the potential hazards associated with their continued use were first reported in 1944 (Davis, 2014). The apparent hazard posed by the use of DDT mainly due to the high levels of dosage, is the ability to alter the functioning of nervous system in humans and domestic animals. This could result in dizziness, convulsions, tremor and instability as a result of tissue poisoning by DDT (Katole *et al.*, 2013). Due to the toxicity and cost associated with the use of DDTs, many people in communal areas have adopted the use of ethnobotanical plants for malaria control. In South Africa, one of the most common of such species is *Lippia javanica* (Lemon bush) which is widely used for its aromatic effect that serves as a repellent for mosquitoes. On the other hand, quantification of plant species used as ethnobotanical plants for controlling *Anopheles* mosquito is necessary for comprehensive approach to malaria mapping and control. This thus provides meaningful contribution to malaria control strategies in the Vhembe District of South Africa. Such contribution is

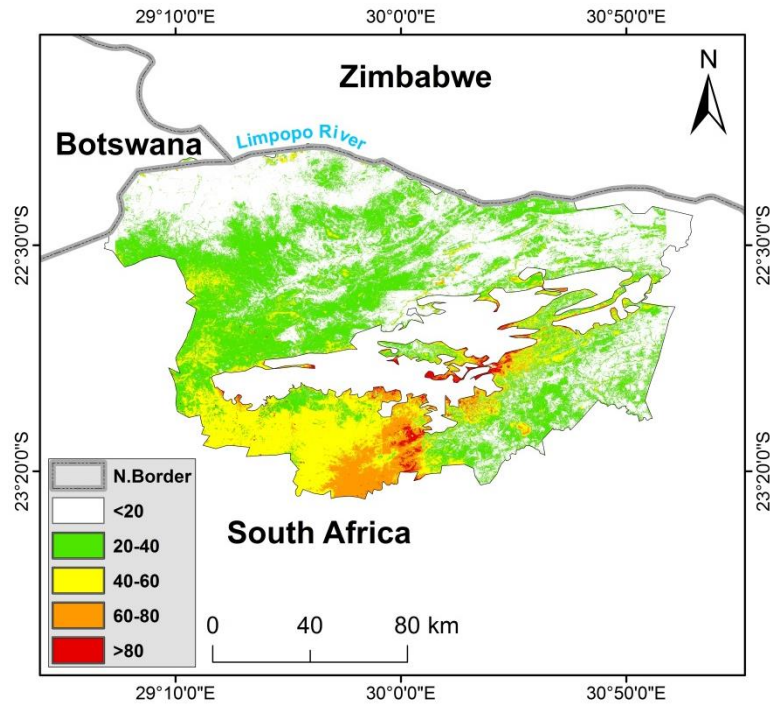
highlighted in three important ways: 1) the improvement of spatial/spectral resolution for species mapping; 2) adding another layer of spatial data onto the already existing topographical layer for defining species habitat and potential ethnobotanical plant use; and 3) ensemble species mapping for a more robust species distribution and habitat preference. In this chapter, we have demonstrated that remote sensing data can aid in predicting the occurrence of *L. javanica* species at 10m spatial resolution. The 10 m spatial resolution allows for more detailed and appropriate species mapping (Valderrama-Landeros *et al.*, 2017), thus contributing to detailed *L. javanica* species inventory and conservation especially when such species is used as an ethnobotanical plant for malaria control (Mabogo, 1990; Maroyi, 2017). Thus the higher spatial resolution *L. javanica* map derived from S2 data is an addition to very coarse species maps that are already available in South Africa at a resolution lower than quarter degree (Foden and Potter, 2005). Both the S2 data and the SRTM data can be easily used to mask environment that is clearly unsuitable for the species of interest (Cord *et al.*, 2013) at high spatial resolutions.

The NDARVI (computed as the target mean reflectance of four red edge bands) and the red edge1 ( $R_{705}$ ) NDVI (also called NDVI<sub>re</sub> in this chapter) were the most significant S2 variables. The significant and positive correlations between these indices and *L. javanica* distribution suggest that the likelihood of finding this plant species increases with the increase in vegetation vigour. The broad-band vegetation indices such as the NDVI, NDWI and the DVI were among the significant S2 variables where the DVI contributing the most in Maxent, while NDWI was the most contributing broadband vegetation index in logistic regression. This indicates that although the narrow-band indices were mostly significant ( $p < 0.001$ ), the broadband indices are also important contributors for explaining variability of species occurrence. The low NDVI and DVI contribution to predictive model shows the inability of the broadband indices to capture subtle vegetation characteristics that define potential micro-habitat of *L. javanica* species. Given the vegetation composition in the study area (dense to sparse vegetation types), the broad-band vegetation indices based on red-NIR reflectance (NDVI, DVI) have suffered from signal saturation (Mutanga and Skidmore, 2004), and this might explain low performance of these indices in individual SDMs. Meanwhile, the narrow-band vegetation indices are known to circumvent saturation problem that results when using broad-band vegetation indices (Baret *et al.*, 1992; Mutanga and Skidmore, 2004; Malahlela *et al.*, 2015). In this chapter, it was clear that the narrow-band red edge vegetation indices were superior to the broad-band indices for characterizing habitat where *L. javanica* occurrence was very likely. However, the correlations between remote sensing variables and the probability of



occurrence of *L. javanica* were fairly low, although significant ( $R^2 < 0.5$ ;  $p < 0.05$ ). The selected remote sensing indices and topographical variables exhibited low correlations with the probable distribution of *L. javanica* species perhaps due to the temporal variation between occurrence data collection (August) and S2 environmental data (August-September). In the VDM, August marks the end of the winter season, the period during which most vegetated landscape enter green-up growth period. On the other hand, September month marks the onset of spring season where savannah vegetation has reached end of peak or maximum growth (Cho *et al.*, 2017). The topographic variables (slope and elevation) derived from SRTM data have shown significant correlation with the distribution of *L. javanica*, thus contributing significantly to predictive accuracies of logistic regression and Maxent Models. Our study showed that slope and Elevation play a major role in determining the species' potential distribution, and this finding is supported by observation made elsewhere by Madzimure *et al.* (2011), who concluded that *L. javanica* was commonly found in grasslands on hillsides in Zimbabwe. In South Africa, a study conducted by Morgenthal *et al.* (2006) found that *L. javanica* sub-community occurred predominantly in habitat located at higher altitudes. On the other hand, the *L. javanica* also occurs in open grassland, in the bush as well as on forest margins (Ng'weno *et al.*, 2009). Although it is thought to be widely distributed on a national scale, in this chapter, the species micro-scale distribution (ecological niche) is rather much narrower. The species distribution in the VDM is more related to elevation, aspect, slope and vegetation greenness which are positively correlated to occurrence of *L. javanica*.

In this chapter, the ensemble SDM has shown to outperform the individual predictive models (logistic and Maxent). This is due to the fact that ensemble modelling combines models that differ in structure, explanatory variables, and data sources, thus allowing inferences that are robust to uncertainties associated with any individual model (Latif *et al.*, 2013; Araujo and New, 2006). The ensemble method was thus used to produce the final species distribution map at 10m spatial resolution. We have demonstrated that the ensemble model for mapping *L. javanica* species which usually occurs as a few stands in various environments (e.g. roadsides, grassland, riverine areas, and bushes). This finding is similar to the one by Marmion *et al.* (2009) who found that the weighted average ensemble model and mean ensemble model provided significantly more robust predictions than all the single-models and the other consensus methods. From the results, it becomes apparent that Maxent outperformed the logistic regression model, although the difference is not profound ( $\pm 2.3\%$ ).



**Figure 6.5.** Ecological niche modelling. Map of the predicted distribution showing the percentage of relative habitat suitability for *L. javanica* species.

By utilizing the weighted average ensemble method, a 10% improvement in the model overall accuracy was achieved over what was achieved through the use of binary logistic regression. However, it should be noted that ensemble methods do not always improve the prediction as their accuracies are largely dependent on the individual models. This means that if the 'input' models have low accuracies and perhaps statistically insignificant predictor variables, the performance of a resultant ensemble model will often be negatively impacted. Additionally, although ensemble approach usually yields better results for species prediction, multiple candidate models for ensemble modelling still require careful development and selection (Latif *et al.*, 2013). Assigning higher weight to Maxent model (0.55 Maxent vs. 0.45 Logistic regression) has resulted in the improvement of prediction capability of SDM, and thus this study advocates for the use of ensemble modelling for species distribution.

The pattern of the predicted species presence at the study area is very important to understanding the environmental factors associated with species recruitment, proliferation and densities. However, the species absence, although largely ignored in SDM's, also provides more complex information about the species and the associated environment, rather than a mere lack of suitability (Lobo 2010). For example, *L. javanica* is largely absent in the western

part of the study area, which is characterized by drier conditions than the areas of predicted species presence. On the other hand, due to dispersal limitations (e.g. ecological barrier such as Soutpansberg Mountain) it is possible that *L. javanica* does not occupy all suitable areas and, as such, caution should be exercised when interpreting the presence results of the predictive maps.

The conservation of *L. javanica* in its natural environment is very essential because of the species relative accessible and cost-effectiveness to the local communities in areas prone to malaria. Moreover, the distribution of *L. javanica* is mostly valuable in areas where western medicine is inaccessibility as a result of high cost associated with purchasing such medicine or where such medicine is unavailable (Light *et al.*, 2005; Mavundaza *et al.*, 2011). Due to its medicinal popularity as a mosquito repellent among the communities living in malaria-prone areas, *L. javanica* occurrence and abundance warrant consistent monitoring since such species forms part of the bigger 'One-Health' approach. Because the analysis in the current study was done from a single S2 image set, the time-series analysis could aid in validating the accuracy of the ensemble modelling results found in this chapter, by providing a multi-temporal variation in vegetation phenology. Moreover, collection of more presence points could aid in more accurate generalization of the predictive models and this could, in turn, improve the performance of the SDM's used in the study. Collecting additional data to increase the sample size is very crucial because the performance of SDM is optimal when the sample size is fairly large, and when species of interest has a narrow niche than a generalist species (Hernandez *et al.*, 2006). The list spectral indices used in the current study is not comprehensive and other Sentinel-derived indices such as those sensitive to chlorophyll may be correlated to the species presence/absence data (Lu *et al.*, 2015; Sonobe and Wang, 2017) and thus shedding light on species detection, monitoring and conservation. Additionally, combining satellite data with environmental data such as soil type, distance to the roads, distance from the rivers or water bodies, shrub patch characteristics, and soil nutrients data could improve the modelling of *L. javanica* species.

## 6.2 Conclusions

In this chapter, the Sentinel-2 vegetation indices have exhibited significant correlation with the distribution of *Lippia javanica* in the Vhembe District. The NDARVI (a normalized index based on average reflectance four red edge bands) was the most significant variable when predicting species distribution using logistic regression, while red edge1 NDVI (NDVI<sub>re</sub>) was the highest

contributing predictor variable in Maxent model. On the other hand, the NDVI, DVI and NDWI have shown correlation with the distribution of *L. javanica* species, at least at micro-scale habitat. The variables derived from SRTM such as elevation, slope and aspect were amongst the most highly contributing variables in all models. The Maxent model performed better than logistic regression (80% and 77.7% respectively), while the combined weighted ensemble model yielded more than 10% improvement to the logistic regression model (86.7% and 77.7% respectively). This study advocates the use of ensemble modelling for mapping suitable habitats for *L. javanica* by taking advantage of high spatial and spectral resolution such as S2 data additional to topographic variables. Mapping the distribution of *L. javanica* in areas prone to malaria is crucial for malaria control particularly in communities that do not have access to western medicine and treatments due to their unavailability or high cost. The new understanding obtained in this chapter also contributes to the update of *L. javanica* distribution map, which has been derived using high spatial resolution dataset. The information will assist the locals in identifying areas suitable for *L. javanica* occurrence, thus enabling ethnobotanical efforts to malaria control.

In summary, this chapter provided a remote sensing approach to delineating areas where *L. javanica* is likely found, based on ground species occurrence data. It has been established in this study, however, that *L. javanica* is more accurately mapped using the combination of SDMs than with the single model. This was done in order in line with objective 4 set out in chapter 1 of this thesis. The use of unequal weighting method is thus encouraged when applying ensemble modelling in ecology and public health.

### 6.3 References

- Adeola A.M., Botai, J.O., Rautenbach H., Adisa O.M., Ncongwane K.P., Botai C.M., Adebayo-Ojo T.C. (2017). Climatic Variables and Malaria Morbidity in Mutale Local Municipality, South Africa: A 19-Year Data Analysis. *International Journal of Environmental Research and Public Health*, 14(11):1360.
- Adjorlolo C., Mutanga O., Cho M.A. (2014). Estimation of canopy nitrogen concentration across C3 and C4 grasslands using Worldview-2 multispectral data. *IEEE Journal Selected Topics in Applied Earth Observation and Remote Sensing*, 7: 4385–4392.



- Álvarez-Martínez J., Jiménez-Alfaro B., Barquín J., Ondiviela B., Recio M., Silió-Calzada A., Juanes J.A. (2017). Modeling the area of occupancy of habitat types with remote sensing. *Methods in Ecology and Evolution*, 9(3): 580 – 593.
- Ansari M.A., Vasudevan P., Tandon M., Razdan R.K. (2000). Larvicidal and mosquito repellent action of peppermint (*Mentha piperita*) oil. *Bioresearch Technology*, 71: 267-271.
- Araujo M.B., New M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, 22(1): 42 – 47.
- Asner G.P., Elmore A.J., Olander L.P., Martin R.E., Harris A.T. (2004). Grazing systems, ecosystem responses, and global change. *Annual Review of Environment and Resources*, 29: 261.
- Asner G.P., Hughes R.F., Vitousek P.M., Knapp D.E., Kennedy-Bowdoin T., Boardman J., Martin R.E., Eastwood M., Green R.O. (2008). Invasive plants transform the three-dimensional structure of rain forests. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11): 4519 – 4523.
- Baldwin R.A. (2009). Use of maximum entropy modelling in wildlife research. *Entropy*, 11:854–866.
- Baret F., Jacquemoud S., Guyot G. (1992). Modeled analysis of the biophysical nature of spectral shift and comparison with information content of broad bands. *Remote Sensing of Environment*, 41:133 – 142.
- Barrell J., Grant J., Hanson A., Mahoney M. (2015). Evaluating the complementarity of acoustic and satellite remote sensing for seagrass landscape mapping. *International Journal of Remote Sensing*, 36(16): 4069 – 4094.
- Bond W.J., Midgley G.F., Woodward F.I. (2002). What controls South African vegetation — climate or fire? *South African Journal of Botany*, 69(1): 79 – 91.
- Bradley B.A., Mustard J.F. (2006). Characterizing the landscape dynamics of an invasive plant and risk of invasion using remote sensing. *Ecological Applications* 16, 1132–1147.
- Breiman L. (2001). Random Forests. *Machine Learning*, 45(1): 5 – 32.
- Chahouki M.A.Z., Sahragard H.P. (2016). Maxent Modelling for Distribution of Plant Species Habitats of Rangelands (Iran). *Polish Journal of Ecology*, 64(4): 453 – 467.
- Chen I., Hill J.K., Ohlemüller R., Roy D.B., Thomas C.D. (2011). Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, 333(6045): 1024 – 1026.

- Cho M.A., Malahlela O., Ramoelo A. (2015). Assessing the utility WorldView-2 imagery for tree species mapping in South African subtropical humid forest and the conservation implications: Dukuduku forest patch as case study. *International Journal of Applied Earth Observation and Geoinformation*, 38: 349–357.
- Cho M.A., Ramoelo A., Dziba L. (2017). Response of Land Surface Phenology to Variation in Tree Cover during Green-Up and Senescence Periods in the Semi-Arid Savanna of Southern Africa. *Remote Sensing*, 9(7): 689. doi:10.3390/rs9070689.
- Clevers J.G.P.W., Kooistra L., van den Brande M.M.M. (2017). Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing*, 9(5): 405. <https://doi.org/10.3390/rs9050405>.
- Collett D. (1991). *Modeling Binary Data*, London: Chapman and Hall.
- Cord A.F., Meentemeyer R.K., Leitao P.J., Vaclavik T. (2013). Modelling species distributions with remote sensing data: bridging disciplinary perspectives. *Journal of Biogeography*, 40: 2226 – 2227.
- Davis F.R. (2014). *Banned: A history of pesticides and the science of toxicology*. Yale University Press. Pp 26. ISBN 978-0300205176. Retrieved 25 October 2018.
- De Oliveria L.F.R., de Oliveira M.L.R., Gomes F.S., Santana R.C. (2017). Estimating foliar nitrogen in Eucalyptus using vegetation indexes. *Scientia Agricola*, 74(2): doi.org/10.1590/1678-992x-2015-0477.
- De Hayes D.H., Jacobson G.L., Schaber P.G., Bongarten B., Iverson L.R., Dieffenbacker-Krall A. (2000). Forest responses to changing climate: lessons from the past and uncertainty for the future. In: Mickler, R.A., Birdsey, R.A., Hom, J.L. (Eds.), *Responses of Northern Forests to Environmental Change*. Springer-Verlag, New York, NY, (ecological studies series), pp. 495–540.
- Drake J.M. (2014). Ensemble algorithms for ecological niche modelling from presence-background and presence-only data. *Ecosphere*, 5(6): 76. <http://dx.doi.org/10.1890/ES13-00202.1>.
- Dudík M., Phillips S.J., Schapire R.E. (2007). Maximum entropy density estimation with generalization regularization and an application to species distribution modelling. *Journal of Machine Learning research*, 8: 1217 – 1260.
- Dudov S.V. (2017). Modeling of species distribution with the use of topography and remote sensing data on the example of vascular plants of the Tukuringra Ridge low mountain belt (Zeya State Nature Reserve, Amur Oblast). *Biology Bulletin Reviews*, 7(3): 246 – 257.



- Edwards T.C., Cutler D.R., Beard K.H., Gibson J. (2007). Predicting invasive plant species occurrences in national parks: a process for prioritizing prevention. Final project report no. 2007-1, USGS Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, UT 84322-5290, USA.
- Elith J., Phillips S.J., Hastie T., Dudik M., Chee Y.E., Yates C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distribution*, 17(1): 43 – 57.
- Foden W., Potter L. (2005). *Lippia javanica* (Burm.f.) Spreng. National Assessment: Red List of South African Plants version 2017.1. Accessed on 2018/03/10. <http://redlist.sanbi.org/species.php?species=4080-3>
- Fourcade Y., Engler J.O., Rodder D., Secondi J. (2014). Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. *PLoS ONE*, 9(5): e97122. <https://doi.org/10.1371/journal.pone.0097122>.
- Frampton W.J., Dash J., Watmough G., Milton E.J (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82: 83 – 92.
- Franklin J (2009) Mapping Species Distributions: Spatial Inference and Prediction. Cambridge, UK: Cambridge University Press.
- Fuller D.O. (2005). Remote detection of invasive *Melaleuca* trees (*Melaleuca quinquenervia*) in South Florida with multispectral IKONOS imagery. *International Journal of Remote Sensing*, 26(5): 1057 – 1063.
- Gamon J.A., Surfus J.S. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytologist*, 143: 105–117.
- Gao B.C. (1996). NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58: 257–266
- Gastón A., García-Viñas J.I. (2011). Modeling species distributions with penalized logistic regressions: A comparison with maximum entropy models. *Ecological Modelling*, 222(13): 2037 – 2041.
- Gillison A.N., Asneer G.P., Fernandes E.C.M., Mafalacusser J., Banze A., Izidine S., da Fonseca A.R., Pacate H. (2016). Biodiversity and agriculture in dynamic landscapes: Integrating ground and remotely-sensed baseline surveys. *Journal of Environmental Management*, 177: 9 – 19.
- Gitelson A., Merzlyak M.N. (1994). Spectral reflectance changes associated 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.A., Merzlyak M.N. (1997). Remote estimation of chlorophyll content in higher plant leaves. *International Journal Remote Sensing*, 18: 2691–2697.
- Gogol-Prokurat, M. (2011). Predicting habitat suitability for rare plants at local spatial scales using a species distribution model. *Ecological Applications*, 21(1): 33–47.
- Gould W. (2000). Remote sensing of vegetation, plant species richness and regional biodiversity hotspots. *Ecological Applications*, 10(6): 1861 – 1870.
- Guillera-Arroita G., Lahoz-Monfort J.J., Elith J., Gordon A., Kujala H., Lentini P.E., McCarthy M.A., Tingley R., Wintle B.A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24: 276 – 292.
- Guisan A., Zimmermann N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135: 147 – 186.
- He K.S., Bradley B.A., Cord A.F., Rocchini D., Tuanmu M., Schmidtlein S., Turner W., Wegmann M., Pettorelli N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, 1(1): 4 – 18.
- He K.S., Rocchini D., Neteller M., Nagendra H. (2011). Benefits of hyperspectral remote sensing for tracking plant invasions. *Biodiversity Review*, 17(3): 381 – 392.
- Hernandez P.A., Graham C.H., Master L.L., Albert D.L. (2006). The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography* 29, 773–785.
- Hill M.J. (2013). Vegetation index suites as indicators of vegetation state in grassland and savanna: An analysis with simulated SENTINEL 2 data for a North American transect. *Remote Sensing of Environment*, 137: 94 – 111.
- Hosmer D.W. Jr, Lemeshow S., Sturdivant R.X. (2013). Applied Logistic Regression. Third Edition. New Jersey: John Wiley & Sons.
- Hosmer D.W., Lemeshow S. (2000). Applied Logistic Regression. Second edition. John Wiley and Sons, Inc., New York.
- Huete A.R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3):295–309.



- Ibáñez I., Silander J.A.Jr., Allen J.M., Treanor S.A., Wilson A. (2009). Identifying hotspots for plant invasions and forecasting focal points of further spread. *Journal of Applied Ecology*, 46: 1219 – 1228.
- Immitzer M., Atzberger C., Koukal T. (2012). Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, 4: 2661 – 2693.
- Ivanescu L., Bodale I., Florescu S., Roman C., Acatrinei D., Miron L. (2016). Climate Change Is Increasing the Risk of the Reemergence of Malaria in Romania. *BioMed Research International*, ID: 8560519. DOI: <http://dx.doi.org/10.1155/2016/8560519> .
- Kabanda T., Munyati C. (2010). Anthropogenic-induced climate change and the resulting tendency to land conflict; The case of the Soutpansberg region, South Africa; Climate Change and Natural Resources Conflicts in Africa. In: Monograph No 170: Institute for Security Studies, Donald Anthony Mwiturubani, Jo-Ansie van Wyk (editors).
- Kempeneers P., Sedano F., Seebach L., Strobl P., San-Miguel-Ayanz J. (2011). Data Fusion of Different Spatial Resolution Remote Sensing Images Applied to Forest-Type Mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12): 4977 – 4986.
- Kganyago M., Odindi J., Adjorlolo C., Mhangara P. (2018). Evaluating the capability of Landsat 8 OLI and SPOT 6 for discriminating invasive alien species in the African Savanna landscape. *International Journal of Applied Earth Observation and Geoinformation*, 67: 10 – 19.
- Khanum R., Mumtaz A.S., Kuma S. (2013). Predicting impacts of climate change on medicinal asclepiads of Pakistan using Maxent modelling. *Acta Oecologica*, 49:23 – 31.
- Korhonen L., Hadi, Packalen P., Rautiainen M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sensing of Environment*, 195: 259 – 274.
- Kumar S., Dash D. (2012). Flora of Nandan Kanan sanctuary: medicinal plants with their role in health care. *International Journal of Pharmacy and Life Sciences*, 3(4): 1631 – 1642.
- Kumar S., Stohlgren T.J. (2009). Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. *Journal of Ecology and the Natural Environment*, 1:94–98.



- Kweka E.J., Mosha F., Lowassa A., Mahande A.M., Kitau J., Matowo J., Mahande M.J., Massenga C.P., Tenu F., Feston E., Lyatuu E.E., Mboya M.A., Mndeme R., Chuwa G., Temu E.A. (2008). Ethnobotanical study of some of mosquito repellent plants in north-eastern Tanzania. *Malaria Journal*, 7(152): <https://doi.org/10.1186/1475-2875-7-152>.
- Laba M., Smith S., Richmond M.E. (2004). Purple loosestrife research and mapping for the Hudson River Valley study area. Final report, New York Cooperative Fish and Wildlife Research Unit, Department of Natural Resources, Cornell University, Ithaca, NY.
- Latif Q.S., Saab V.A., Dudley J.G., Hollenbeck J.P. (2013). Ensemble modeling to predict habitat suitability for a large-scale disturbance specialist. *Ecology and Evolution*, 3(13):4348 – 4364.
- Lecain D.R., Morgan J.A., Schuman G.E., Reeder J.D., Hart R.H. (2000). Carbon exchange rates in grazed and ungrazed pastures of Wyoming. *Journal of Rangeland Management*, 53:199-206
- Legendre P., Dale M.R.T., Fortin M., Gurevitch J., Hohn M., Myers D. (2002). Consequences of structure for the design and analysis of ecological field surveys. *Ecography*, 25: 601 – 615.
- Lehmann C.E.R., Archibald S.A., Hoffmann W.A., Bond W.J. (2011). Deciphering the distribution of the savanna biome. *New Phytologist*, 191: 197 – 209.
- Lemke D., Brown J.A. (2012). Habitat modelling of alien plant species at varying levels of occupancy. *Forests*, 3:799 – 817.
- Lengeler, C. (2004) Insecticide-treated bed nets and curtains for preventing malaria. Cochrane Database System Reviews 2, CD000363.
- Light M.E., Sparg S.G., Stafford G.I., Van Staden J. (2005). Riding the wave: South Africa's contribution to ethnopharmacological research over the last 25 years. *Journal of Ethnopharmacology*, 100: 127–130
- Lu S., Lu X., Zhao W., Liu Y., Wang Z., Omasa K. (2015). Comparing vegetation indices for remote chlorophyll measurement of white poplar and Chinese elm leaves with different adaxial and abaxial surfaces. *Journal of Experimental Botany*, 66(18): 5625 – 5637.
- Lukwa N. (1994). Do traditional mosquito repellent plants work as mosquito larvicides? *Central African Journal of Medicine*, 40(11): 306-9.



- Mabogo D.E.N. (1990). The Ethnobotany of the Vhavenda. M.Sc. Thesis. University of Pretoria.
- Madonsela S., Cho M.A., Ramoelo A., Mutanga O., Naidoo L. (2018). Estimating tree species diversity in the savannah using NDVI and woody canopy cover. *International Journal of Applied Earth Observation and Geoinformation*, 66: 106 – 115.
- Madzimure J., Nyahangare E.T., Hamudikuwanda H., Hove T., Stevenson P.C., Belmain S.R., Mvumi B.M. (2011). Acaricidal efficacy against cattle ticks and acute oral toxicity of *Lippia javanica* (Burm F.) Spreng. *Tropical Animal Health and Production*, 43: 481 – 489.
- Maharaj R., Morris N., Seocharan I., Kruger P., Moonasar D., Mabuza A., Raswiswi E., Raman J. (2012). The feasibility of malaria elimination in South Africa. *Malaria Journal*, 11:423. DOI: 10.1186/1475-2875-11-423.
- Maheu-Giroux M., de Blois S. (2004). Mapping the invasive species *Phragmites australis* in linear wetland corridors. *Aquatic Botany* 83:310–320.
- Malahlela O.E., Cho M.A., Mutanga O. (2015). Mapping the occurrence of *Chromolaena odorata* (L.) in subtropical forest gaps using environmental and remote sensing data. *Biological Invasions*, 17(7): 2027–2042.
- Malahlela O.E., Olwoch J.M., Adjorlolo C. (2018). Evaluating Efficacy of Landsat-Derived Environmental Covariates for Predicting Malaria Distribution in Rural Villages of Vhembe District, South Africa. *EcoHealth*, 15(1): 23 – 40.
- Marmion M., Parviainen M., Luoto M., Heikkinen R.K., Thuiller W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distribution*, 15: 59 – 69.
- Maroyi A. (2017). *Lippia javanica* (Burm.f.) Spreng.: Traditional and Commercial Uses and Phytochemical and Pharmacological Significance in the African and Indian Subcontinent. *Evidence-Based Complementary and Alternative Medicine*, 2017: ID.6746071.
- Martin T.G., Wintle B.A., Rhodes J.R., Kuhnert P.M., Field S.A., Low-Choy S.J., Tyre A.J., Possingham H.P. (2005). Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8(11): 1235-1246.
- Matthies D., Bräuer I., Maibom W., Tschardt T. (2004). Population size and the risk of local extinction: empirical evidence from rare plants. *Oikos*, 105(3):481–8.



- Mavundza E.J., Maharaj R., Finnie J.F., Kabera G., Van Staden J. (2011). An ethnobotanical survey of mosquito repellent plants in uMkhanyakude district, KwaZulu-Natal province, South Africa. *Journal of Ethnopharmacology*, 137(3):1516 – 1520.
- Morgenthal T.L., Kellner K., van Rensburg L., van der Merwe J.P.A. (2006). Vegetation and habitat types of the Umkhanyakud Node. *South African Journal of Botany*, 72(1): 1 – 10.
- Mpandeli S. (2014). Managing climate risks using seasonal climate forecast information in Vhembe District in Limpopo Province, South Africa. *Journal of Sustainable Development*, 7(5):68–81.
- Müllerová J., Pergl J., Pysek P. (2013). Remote sensing as a tool for monitoring plant invasions: Testing the effects of data resolution and image classification approach on the detection of a model plant species *Heracleum mantegazzianum* (giant hogweed).
- Müllerová J., Pysek P., Jarosík V., Pergl J. (2005). Aerial photographs as a tool for assessing the regional dynamics of the invasive plant species *Heracleum mantegazzianum*. *Journal of Applied Ecology*, 42: 1–12.
- Müller-Wilm U. (2016). Sentinel-2 MSI – Level-2A Prototype Processor Installation and User Manual. S2PAD-VEGA-SUM-0001, Issue (2.2).
- Muñoz J., Feleccísimo Á.M. (2004). Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science*, 15: 285 – 292.
- Mutanga O., Skidmore A.K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19): 3999 – 4014.
- Nagendra H., Rocchini D., Ghate R., Sharma B., Pareeth S. (2010). Assessing plant diversity in a dry tropical forest: comparing the utility of Landsat and IKONOS satellite images. *Remote Sensing*, 2: 478–496.
- Ng'weno C.C., Mwasi S.M., Kairu J.K. (2009). Distribution, density and impact of invasive plants in Lake Nakuru National Park, Kenya. *African Journal of Ecology*, 48: 905–913.
- Padalia H., Bharti R.R., Pundir Y.P.S., Sharma K.P. (2010). Geospatial multiple logistic regression approach for habitat characterization of scarce plant population: A case study of *Pittosporum eriocarpum* Royle (an endemic species of Uttarakhand, India). *Journal of the Indian Society of Remote Sensing*, 38(3): 513 – 521.
- Pau S., Gillespie T.W., Wolkovich E.M. (2012). Dissecting NDVI – species richness relationships in Hawaiian dry forests. *Journal of Biogeography*, 39(9): 1678 – 1686.

- Peerbhay K., Mutanga O., Lottering R., Ismail R. (2016). Mapping *Solanum mauritianum* plant invasions using WorldView-2 imagery and unsupervised random forests. *Remote Sensing of Environment*, 182: 39 – 48.
- Peng Y., Fan M., Song J., Cui T., Li R. (2018). Assessment of plant species diversity based on hyperspectral indices at a fine scale. *Nature Scientific Reports*, 8:4776. doi:10.1038/s41598-018-23136-5.
- Peterson E.B. (2005). Estimating cover of an invasive grass *Bromus tectorum* using logistic regression and phenology derived from two dates of Landsat ETM+ data. *International Journal of Remote Sensing*, 26: 2491–2507.
- Phaswane M.C., Masevhe N.A. (2018). Investigation of medicinal plants for management of malaria at Tshakhuma community, Makhado Municipality, Vhembe District, Limpopo province, Republic of South Africa. *South African Journal of Botany*, 115: 327
- Phillips S.J., Anderston R.P., Schapire R.E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190(3-4): 231 – 259.
- Pierre-Louis A-M, Qamruddin J, Espinosa I, Challa S (2018). The Malaria Control Success Story. Chapter 23: 417-432. World Bank.
- Pollnac F., Seipel T., Repath C., Rew L.J. (2012). Plant invasion at landscape and local scales along roadways in the mountainous region of the Greater Yellowstone Ecosystem. *Biological Invasions*, 14:1753–1763.
- Pulliam H. (2000). On the relationship between niche and distribution. *Ecology Letters*, 3: 349 – 361.
- QGIS Development Team (2018). QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org> .
- Ramoelo A., Cho M.A. (2018). Explaining Leaf Nitrogen Distribution in a Semi-Arid Environment Predicted on Sentinel-2 Imagery Using a Field Spectroscopy Derived Model. *Remote Sensing*, 10(2): 269. doi:10.3390/rs10020269.
- Ramoelo A., Cho M.A., Mathieu R., Madonsela S., van de Kerchove R., Kaszta Z., Wolff E. (2015). Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 43: 43 – 54.
- Ramoelo A., Dzikiti S., van Deventer H., Maherry A., Cho M.A., Gush M. (2015). Potential to monitor plant stress using remote sensing tools. *Journal of Arid Environments*, 113: 134 – 144.





- Rocchini D., Balkenhol N., Carter G.A., Foody G.M., Gillespie T.W., He K.S., Kark S., Levin N., Lucas K., Luoto M., Nagendra H., Oldeland J., Ricotta C., Southworth J., Neteler M. (2010). Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecological Informatics*, 5(5): 318 – 329.
- Rondeaux G., Steven M., Baret F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55: 95–107.
- Ross L.C., Lambdom P.W., Hume P.E. (2008). Disentangling the roles of climate, propagule pressure, and land use on the current and potential elevation distribution of the invasive weed *Oxalis pes-caprae* L. on Crete. *Perspectives in Plant Ecology, Evolution, and Systematics*, 10: 251 – 258.
- Rouse J.W., Haas R.H., Schell J.A., Deering D.W., Harlan J.C. (1974). Monitoring the Vernal Advancements and Retrogradation (Greenwave Effect) of Nature Vegetation; NASA/GSFC Final Report; NASA: Greenbelt, MD, USA.
- Sahragard P.H., Ajourlo M. (2018). A comparison of logistic regression and maximum entropy for distribution modeling of range plant species (a case study in rangelands of western Taftan, southeastern Iran). *Turkish Journal of Botany*, 42: 28 – 37.
- Samie A., Obi C.L., Bessong P.O., Namrita L. (2005). Activity profiles of fourteen selected medicinal plants from Rural Venda communities in South Africa against fifteen clinical bacterial species. *African Journal of Biotechnology*, 4 (12):1443-1451.
- Sharma V.P. (1996). Re-emergence of malaria in India. *Indian Journal of Medical Research*, 103: 26 – 45
- Schumacher P., Mislinsheva B., Brenning A., Zandler H., Brandt M., Samimi C., Koellner T. (2016). Do Red Edge and Texture Attributes from High-Resolution Satellite Data Improve Wood Volume Estimation in a Semi-Arid Mountainous Region? *Remote Sensing*, 8(7): 540.
- Seyoum A., Pålsson K., Kung'a S., Kabiru E.W., Lwande W., Killeen W.F., Hassanali A., Knots B.G.J. (2002). Traditional use of mosquito-repellent plants in western Kenya and their evaluation in semi-field experimental huts against *Anopheles gambiae*: ethnobotanical studies and application by thermal expulsion and direct burning. *Journal of Urology*, 96(3): 225 – 231.
- Sharma V.P. (1996). Re-emergence of malaria in India. *Indian Journal of Medical Research*, 103: 26 – 45





- Smith A., Page K., Duffy K., Slotow R. (2012). Using Maximum Entropy modeling to predict the potential distributions of large trees for conservation planning. *Ecosphere*, 3(6):56. <http://dx.doi.org/10.1890/ES12-00053.1>.
- Sonobe R., Wang Q. (2017). Towards a universal hyperspectral index to assess chlorophyll content in deciduous forests. *Remote Sensing*, 9(191): doi:10.3390/rs9030191.
- South African Weather Service (SAWS) (2018). What are the temperature, rainfall and wind extremes in SA? <http://www.weathersa.co.za/learning/climate-questions/39-what-are-the-temperature-rainfall-and-wind-extremes-in-sa> . Accessed: 21/04/2018.
- Statistics South Africa. (2012). Population census 2011. <http://www.statssa.gov.za/publications/SASStatistics/SASStatistics2012.pdf>
- Statistics South Africa. (2016). Population census 2016. Vhembe District Municipality. [http://cs2016.statssa.gov.za/?page\\_id=270](http://cs2016.statssa.gov.za/?page_id=270).
- Stohlgren T.J., Ma P., Kumar S., Rocca M., Morisette J.T., Jarnevich C.S., Benson N. (2010). Ensemble habitat mapping of invasive plant species. *International Journal of Risk Analysis*, 30(2): 224 – 235.
- Tucker C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8: 127–150.
- Ullerud H.A., Bryn A., Klanderud K. (2016). Distribution modelling of vegetation types in the boreal–alpine ecotone. *Applied Vegetation Science*, 19: 528 – 540.
- Ustin S.L., Gitelson A.A., Jacquemoud S., Schaepman M., Asner G.P., Gamon J.A., Zarco-Tejada P.(2009). Retrieval of Foliar Information about Plant Pigment Systems from High Resolution Spectroscopy. *Remote Sensing of Environment*, 113 (Suppl. 1): S67–S77. DOI:10.1016/j.rse.2008.10.019.
- Valderrama-Landeros L., Flores-de-Santiago F., Kovacs J.M., Flores-Verdugo F. (2017). An assessment of commonly employed satellite-based remote sensors for mapping mangrove species in Mexico using an NDVI-based classification scheme. *Environmental Monitoring and Assessment*, 190(1):23. DOI: 10.1007/s10661-017-6399-z.
- Van Deventer H., Cho M.A., Mutanga O., Ramoelo A. (2015). Capability of models to predict leaf n and p across four seasons for six sub-tropical forest evergreen trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101: 209–220



- Viljoen A.M., Subramoney S., van Vuuren S.F., Bas, er K.H.C., Demirci B. (2005). The composition, geographical variation and antimicrobial activity of *Lippia javanica* (Verbenaceae) leaf essential oils. *Journal of Ethnopharmacology*, 96:271–277.
- Wakie T.T., Evangelista P.H., Jarnevich C.S., Laituri M. (2014). Mapping Current and Potential Distribution of Non-Native *Prosopis juliflora* in the Afar Region of Ethiopia. *PLoS ONE*, 9(11): e112854. DOI:10.1371/journal.pone.0112854.
- Wang H., Wonkka C.L., Treglia M.L., Grant W.E., Smeins F.E., Rogers W.E. (2015). of an endangered endemic orchid. *AoB Plants*, 7:plv039. doi:10.1093/aobpla/plv039.
- West A.M., Evangelista P.H., Jarnevich C.S., Young N.E., Stohlgren T.J., Talbert C., Talbert M., Morisette J., Anderson R. (2016a). Integrating remote sensing with species distribution models; mapping Tamarisk invasions using the Software for Assisted Habitat Modeling (SAHM). *Journal of Visualized Experiments*, 116:54578. DOI: 10.3791/54578.
- West A.M., Kumar S., Brown C.S., Stohlgren T.J., Bromberg J. (2016b). Field validation of an invasive species Maxent model. *Ecological Informatics*, 36: 126 – 134.
- WHO (2008) World Malaria Report 2008, WHO.
- World Health Organization (WHO). (2018). Update on the E-2020 Initiative of 21 Malaria-Elimination Countries: *Report and Country Briefs*. WHO Geneva, Switzerland.
- Yang X., Skidmore A.K., Melick D.R., Zhou Z., XU J. (2006). Mapping non-wood forest product (Matsutake mushrooms) using logistic regression and a GIS expert system. *Ecological Modelling*, 19(1 – 2): 208 – 218.
- Zarco-Tejada P.J., Hornero A., Hernandez-Clemente R., Beck P.S.A. (2018). Understanding the temporal dimension of the red-edge spectral region for forest decline detection using high-resolution hyperspectral and Sentinel-2a imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 137: 134 – 148.
- Zeng Y., Wang S., Luan J., Yu M. (2017). The Analysis of Tree Species Distribution Information Extraction and Landscape Pattern Based on Remote Sensing Images. *Frontiers in Environmental Science*, 5:46. <https://doi.org/10.3389/fenvs.2017.00046>.
- Zhu Y, Liu K., Myint S.W., Liu H., He Z. (2017). Exploring the Potential of WorldView-2 Red-Edge Band-Based Vegetation Indices for Estimation of Mangrove Leaf Area Index with Machine Learning Algorithms. *Remote Sensing*, 9(1060): doidoi:10.3390/rs9101060



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA



## **CHAPTER 7:**

Remote sensing of environmental variables for mapping  
malaria distribution: the case of Vhembe District  
Municipality, South Africa

Summary and conclusions

## 7.1 Introduction

Previous research findings have indicated that mapping malaria at local and regional scales is crucial for effective disease management and control (Dalrymple *et al.*, 2015; Okami and Kohtake, 2017; Gwitira *et al.*, 2018). Similarly, findings from this research also support the need to map malaria occurrence and factors that contribute significantly to malaria transmission in malaria endemic areas. The main contribution of malaria mapping study is mainly to inform the targeted malaria intervention by health professionals and to contribute to the baseline data for early-warning systems. The potential habitats of malaria vectors have long been associated with vegetation greenness and water bodies. However, there is a need to accurately quantify the potential habitats of malaria vectors for informed decision-making with regard to control of malaria incidences in the subtropical regions of the world. By so doing, costs associated with disease control and elimination can be minimized while improving the inventory of the resources required for controlling outdoor malaria incidence rates. Remote sensing has offered avenues for cost-effective and rapid characterization the biophysical and biochemical parameters of outdoor habitats for *Anopheline* mosquitoes, particularly in rural areas with rugged terrain (Malahlela *et al.*, 2018).

Although the capabilities of remote sensing for mapping infectious diseases have previously been demonstrated (Adeola *et al.*, 2016), there are still some challenges that are related to spatial and spectral configuration of most conventional optical sensors. The spatial resolution of the conventional satellites commonly used in public health studies is often too coarse to capture the biophysical and biochemical parameters of malaria vector habitats at a local scale. The ability of EO satellites to capture the local scale environmental conditions is very crucial in vector biology and diseases transmission (Richards *et al.*, 2010; Minakwa *et al.*, 2002), because local environmental conditions associated with micro-climate also contribution to survival of malaria pathogens (Paaijmans and Thomas, 2011). Additionally, the limited ability to discriminate amongst biophysical and biochemical properties related to malaria vector habitats through conventional satellite spectral bands can be a handicap to the use of traditional satellite datasets for malaria mapping. In order to circumvent this challenge, the use of hyperspectral bands was recommended by many researchers due to their ability to capture subtle variations in habitat biophysical and biochemical characteristics (Mutanga and Skidmore, 2004). Although the hyperspectral remote sensing has been successfully used for soil (Peon *et al.*, 2017), water (Garcia *et al.*, 2018), and vegetation (Hirano *et al.*, 2003) studies, it is often marred with high data dimensionality problem that introduces multicollinearity and demands high computational power (Adjorlolo *et al.*, 2013; Rajan *et al.*, 2008). Apart from these apparent disadvantages, previous research has shown that both hyperspectral and

multispectral sensors have unique advantages that, when well-utilized, complement each other. One of the fundamental questions asked in this thesis was whether it is possible to characterize malaria vector habitats from optical remote sensing dataset that has the advantage of both the multispectral and hyperspectral sensors and at high spatial resolution? The aim of this research was to characterize the remote sensing variables necessary to map the spatial distribution of malaria in the Vhembe District Municipality. The objectives were mainly focused on (i) characterizing *Anopheles* mosquito habitats, (ii) mapping the spatial distribution of malaria pathogen in Vhembe, and (iii) delineating the occurrence of *L. javanica* for the localized malaria control strategy.

The challenges and opportunities related to the use of remote sensing for infectious diseases mapping are reviewed, with specific reference to malaria.

### **7.1.1 Challenges and opportunities of remote sensing for malaria mapping**

Mapping the potential habitats for malaria vectors is one of the challenging steps for malaria control strategies. Since the realization of the contribution of vegetation, soil and water to malaria incidence rates, attempts have been made to relate such habitats with remote sensing approaches. However, from the literature review it becomes evident that most of the studies that utilized remote sensing relied on the indication given by vegetation greenness indices (NDVI, for example) while neglected some vegetation parameters which are very closely linked to mosquito's resting/questing behaviour. The high resolution spatial information on the distribution of mosquito micro-habitats such as cattle footprints and puddles (natural or man-made), which are important for malaria transmission is still missing, to a greater extent. On the other hand, characterizing vegetation parameters that largely dictate the survival of adult mosquitoes (e.g. LAI) in relation to malaria has not been adequately done, and as such more work is required to assess the potential use of high resolution satellite dataset for estimating LAI in areas endemic to malaria.

The improvement in radiometric, spectral and spatio-temporal configuration of current and future EO satellites offers new opportunities for improved quantification of *Anopheles* habitats. This thus makes it possible to untangle various biophysical and biochemical characteristics of mosquito breeding (water) and resting (vegetation) sites that was, otherwise, difficult to quantify using traditional remote sensing datasets. For example, the presence of narrow-band

such as the red edge in WorldView-2, RapidEye, and Sentinel-2 offers opportunities to characterize variations in physico-chemical constitution of soil-water-vegetation nexus. For example, [Kross \*et al.\* \(2015\)](#) found that the vegetation indices computed from the red-edge band performed consistently better than the traditional red band indices when estimation vegetation leaf area index (LAI). However, traditional classification methods designed for conventional datasets are unable to produce consistent and robust results when mapping malaria, due to the high data dimensionality of EO datasets such as the hyperspectral data. The improvements in sensor specifications also demand rapid algorithms geared towards handling and processing high data volumes. In summary the following conclusions were made during the literature review (Chapter 2):

- The distribution and abundance of malaria vectors in tropical regions are highly correlated to the occurrence of micro-habitats such as cattle hoofprints, puddles and warm, moist vegetated areas.
- The survival of adult *Anopheles* mosquito is dependent upon conducive environmental conditions that include the architecture of vegetation canopy, with varying incoming radiation attenuation characteristics.
- The use of reflectance properties of water, soil and vegetation could enhance mapping of malaria at high resolution in malaria endemic areas.

### **7.1.2 Estimating the distribution of cattle hoofprints**

Cattle hoofprints are some of the most crucial habitats for *An.arabiensis* and *An.fenestus* in Southern Africa and their distribution was modelled using remote sensing data and field data with the geostatistical methods. Figure 7.1 shows the results of the cattle hoofprint modelling from Sentinel-2 dataset. The results show that the co-kriging method that takes into account narrow-band remote sensing variables is better correlated with the distribution of cattle hoofprints than both the ordinary kriging and the regression methods (Table 7.2). The implication of these results testifies that narrow-band remote sensing bands such as the red edge present in Sentinel-2 offers improved capability of mapping malaria vector habitats which



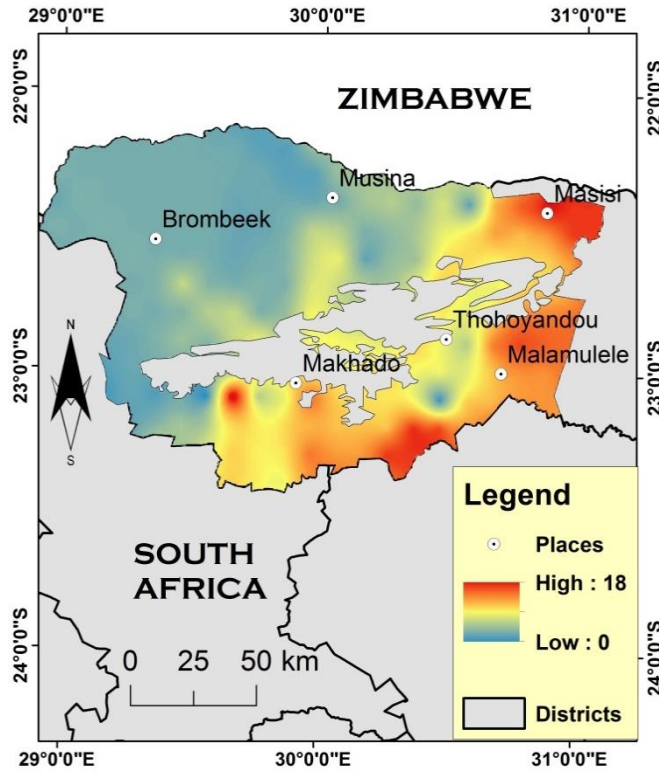


Figure 7.1. Map that shows the distribution of cattle hoofprints per 100 m<sup>2</sup> in the Vhembe District Municipality

**Table 7.1:**

Comparison of interpolation performance among OK, CK and SMLR for predicting cattle hoofprints

Method	Validation		
	$R^2$	RMSE	MAD
OK	0.57	2.39	2.11
CK	0.69	0.20	0.04
SMLR	0.25	5.20	4.55

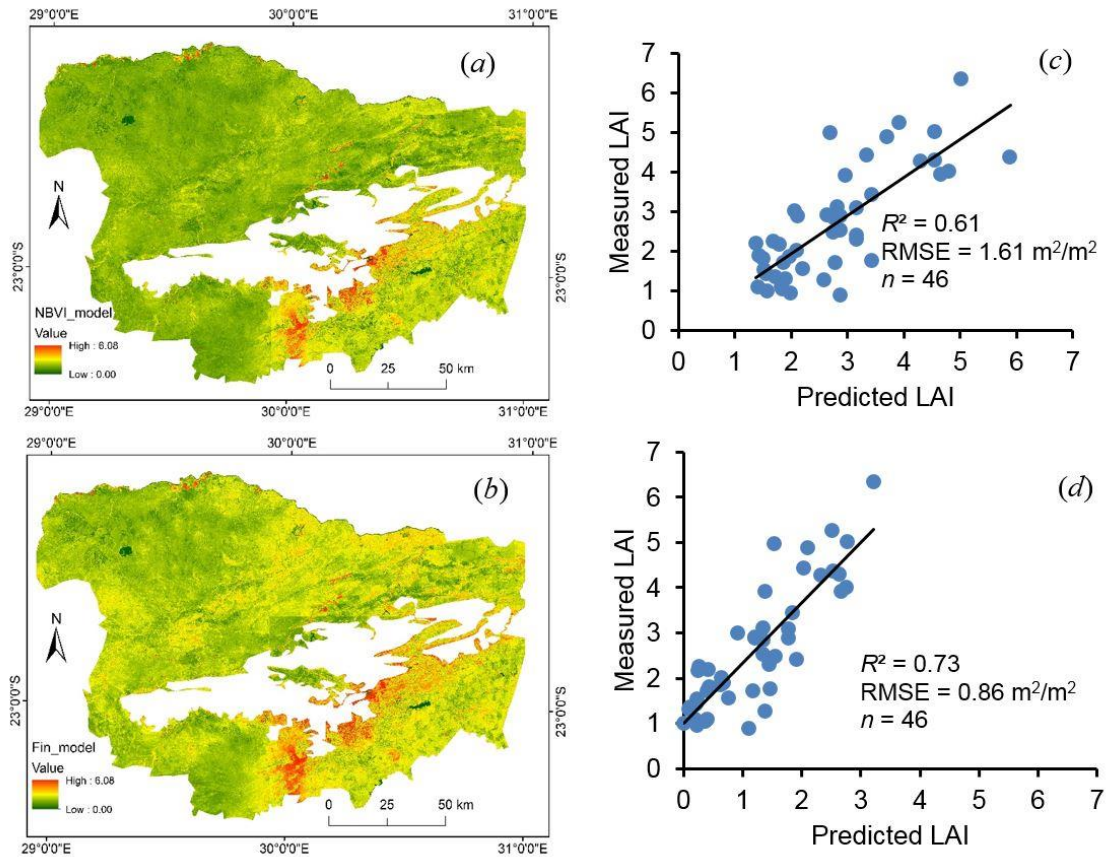
$R^2$  = coefficient of determination, RMSE = root mean square error, MAD = mean absolute deviance, OK = ordinary kriging, CK = co-kriging, SMLR = stepwise multiple linear regression.

In summary, the following conclusive remarks were drawn from the findings presented in Chapter 3:

- The high resolution Sentinel-2 data was correlated to the distribution of cattle hoofprints. The general trend was found to be significant correlations between blue, red edge and Ferric Iron Index derived from Sentinel-2 imagery.
- High predictive accuracy was achieved through co-kriging of field observation footprint dataset and the Sentinel-2 data ( $R^2 = 0.69$ ). In addition the co-kriging method resulted in low predictive errors, with RMSE = 0.2 prints per 100 m<sup>2</sup>; MAD = 0.04 prints per 100 m<sup>2</sup>) using independent validation dataset.
- Most of the cattle hoofprints are found in the north-eastern (Masisi area), eastern (Malamulele/Makuleke area) and south central parts (Thohoyandou to Makhado areas) of the Vhembe District Municipality (Figure 7.1). These areas are known to comprise of high malaria incidence rates in the VDM, most of which are situated in the Mutale local municipality (Khosa *et al.*, 2013). Lower predictions were made for areas in the northern and western parts of the study area.
- The high resolution satellite data with both multispectral and hyperspectral configuration offers new opportunities for malaria vector habitat mapping.

### **7.1.3 Characterizing mosquito resting and questing habitats (LAI)**

This section (Chapter 4) tested the performance of broadband vegetation indices (BBVI) and the narrowband vegetation (NBVI) indices for retrieval of leaf area index (LAI) calibrated from heterogeneous landscape in the malaria prone environments. The results from this chapter indicate that the BBVI called modified chlorophyll absorption ratio index (MCARI<sub>2</sub>) and the modified triangular vegetation index (MTVI<sub>2</sub>) have shown higher correlation with the LAI distribution ( $R^2 = 0.73$ ; RMSE = 0.86 m<sup>2</sup> m<sup>-2</sup>) than the NBVI computed from the red edge region ( $R^2 = 0.61$ ; RMSE = 1.61 m<sup>2</sup> m<sup>-2</sup>). This chapter emphasizes the importance of the Sentinel-2 broad bands (visible to near-infrared), acquired at 10 m spatial resolution which is the important finding in this chapter. Findings from this chapter indicate a robust generalization of LAI estimation, because the parameterization was made from landscape with heterogeneous vegetation communities. Figure 7.2 shows the results of the models used for mapping LAI at the study area.



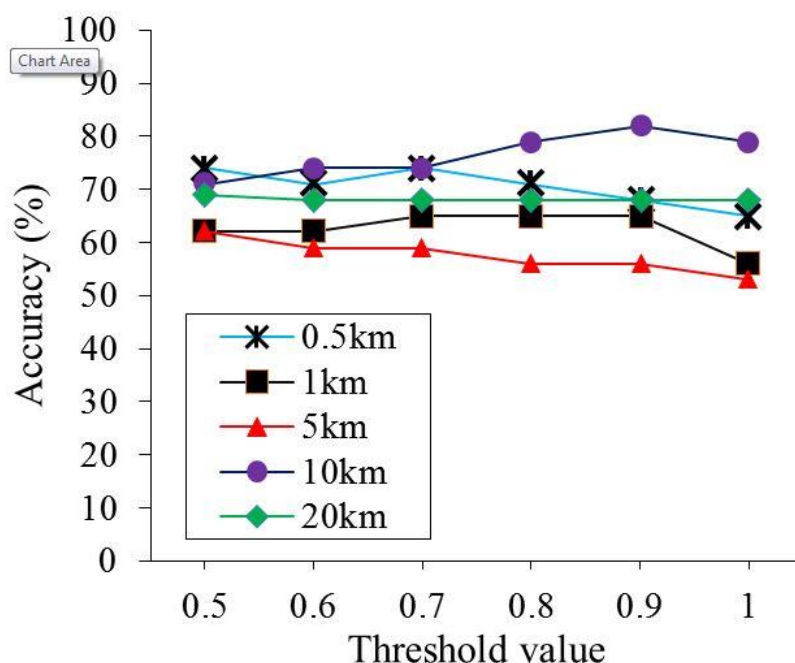
**Figure 7.2:** The results of LAI retrieval from narrow-band vegetation indices (a) vs. the final broadband vegetation indices (b) at the study area. The scatterplots show the respective correlation between measured and predicted LAI for the NBVI (c) and the final BBVI model (d) respectively.

From the current chapter the following conclusions were made:

- The broadband vegetation indices have shown superior performance for retrieving LAI in the study area than the narrowband indices computed from the red edge band of Sentinel-2.
- In general, the retrieval accuracy of both the BBVI and NBVI yielded the correlation of above 60% with the validation dataset ( $n = 46$ ).
- Sentinel-2 offers the potential to retrieve the biophysical and biochemical characteristics of vegetation (i.e. LAI) which is an indicator of where mosquitoes tend to rest during the day in high malaria transmission season.

#### 7.1.4 Mapping the distribution of malaria in the study area

The prediction of malaria is central to understanding the burden of disease and the implementation of intervention measures on disease hot spots. This study (Chapter 5) has shown that remote sensing indices that are sensitive to vegetation moisture and vegetation greenness are better predictors of malaria distribution when using presence and (pseudo) absence *Plasmodium falciparum* generated at various distances from the presence location. The pseudo-absences generated at 10 km from the known have yielded higher accuracies than those generated at 0.5 km, 1 km, 5 km, and 20 km.



**Figure 7.3.** The accuracy of malaria mapping when pseudo-absences are drawn from 0.5, 1, 5, 10 and 20 km from the known presence locations.

Due to the highest mapping accuracy obtained when the pseudo-absences were drawn at 10 km from the known presence locations, the following conclusions were drawn:

- The study has shown that suitable habitats of malaria vectors are generally found within a radius of 10km in semi-arid environments and this insight can be useful to aid efforts aimed at putting in place evidence based preventative measures against malaria infections.
- Furthermore, this result is important in understanding malaria dynamics under the current climate and environmental changes.
- The study has also demonstrated the use of conventional spectral such as those of Landsat Thematic Mapper and the ability to extract environmental conditions which

favour the distribution of malaria vector (*An. arabiensis*) such as the canopy moisture content in vegetation, which serves as a surrogate for rainfall.

### **7.1.5 Mapping the spatial distribution of *L. javanica* for malaria control**

What is the significance of mapping malaria disease distribution and factors affecting vector habitats but not providing spatial information on the existing mechanisms for disease control? Many countries, especially in Africa have developed their own malaria control plans and subsequent implementations resulted in the decrease in the number of malaria transmissions (Pierre-Louis *et al.*, 2018). For example, many of these countries have adopted the use of insecticide treated nets (ITNs) and indoor residual sprays (IRS), which have become cornerstones for malaria control programme (WHO, 2008; Lengelar, 2004). In most African communities, the use of repellents of plant origin has been one of the most effective methods for controlling malaria for many years (Kweka *et al.*, 2008), because of their cheaper costs of treatment where such plant species are commonly found, and have shown to contain low toxicity to humans and animals (Ansari *et al.*, 2000). Many plant species have been effective alternative for repelling malaria vectors, and these plant species include *Lantana camara*, *Omicum americanum*, *Azadirachta indica*, and *Lippia javanica* (Seyoum *et al.*, 2002; Mabogo, 1990).

Remote sensing offers fast and efficient alternative approach for mapping the potential distribution of plant species. This is because remote sensing data is acquired for much larger areas at various spatio-temporal and spectral resolutions (Kempeneers *et al.*, 2011). On the other hand, the spectral configuration of remote sensing instruments allows for discrimination amongst various plant species and to characterize topographical and environmental features that are associated with habitat of species of interest (Peng *et al.*, 2018; Álvarez-Martínez *et al.*, 2017). In this study (Chapter 6), Sentinel-2 data with 10 multispectral bands (and 10m spatial resolution) was tested for mapping the distribution of *L. javanica* species commonly used for controlling malaria vectors at the study area.

The study utilized the ensemble species modelling approach that included Maxent and stepwise logistic regression models for predicting species distribution. Table 7.2 shows the performance of three models for mapping species distribution.

**Table 7.2:**

Predicted occurrence vs. observed occurrence of *L. javanica* ( $n = 45$ ) using three SDM's at probability threshold  $> 0.6$ .

		presence	absence	total
<b>Logistic</b>	presence	1	2	3
	absence	8	34	42
<b>Maxent</b>		presence	absence	
	presence	1	0	1
	absence	9	35	44
<b>Ensemble</b>		presence	absence	
	presence	3	0	3
	absence	6	36	42

In summary:

- The NDARVI (a normalized index based on average reflectance four red edge bands) was the most significant variable when predicting species distribution using logistic regression, while red edge1 NDVI (NDVI<sub>re</sub>) was the highest contributing predictor variable in Maxent model.
- The variables derived from SRTM such as elevation, slope and aspect were amongst the most highly contributing variables in all models.
- The Maxent model performed better than logistic regression (80% and 77.7% respectively), while the combined weighted ensemble model yielded more than 10% improvement to the logistic regression model (86.7% and 77.7% respectively).
- The new understanding obtained in this study also contributes to the update of *L. javanica* distribution map, which has been derived using high spatial resolution dataset. The information will assist the locals in identifying areas suitable for *L. javanica* occurrence, thus enabling ethnobotanical efforts to malaria control.
- The mapping of *L. javanica* enhances the indigenous knowledge system (IKS) of the local community in the Vhembe District (mostly the Vha-Venda people) concerning the locations of plant species that is usually used for repelling mosquitoes. The space and space-related technologies such as the one adopted for this study thus offer high resolution species mapping which can be communicated to local communities who continue to practice malaria preventative medicine based on IKS.



## 7.2 Overall conclusions and future work

This study contributed significantly to the application of remote sensing for epidemiological research and public health. Findings from this thesis show that remote sensing indices designed for detecting changes in vegetation greenness, moisture content, and soil bareness can be used to derive information for mapping the potential breeding and resting sites of *Anopheles arabiensis* and *An.fenestus* in the subtropical environment. The success of mapping these parameters relies strongly on the establishment of spectral relationships between disease distribution, malaria vector habitat's biophysical and biochemical characteristics and remote sensing datasets. The major finding from this study lies in the need to characterize habitats of *Anopheline* mosquito complex in Southern African tropical zones by employing remote sensing techniques. It was thus demonstrated that the high resolution satellite data has the potential to characterize: (i) the potential breeding sites for the malaria vectors in southern Africa, which are usually formed from cattle hoofprints, (ii) the outdoor resting and questing sites (refuge) for the malaria vectors, which tend to take refuge under the plant leaves to avoid predators and desiccation, (iii) favourable environmental conditions associated with vegetation biophysical and biochemical parameters necessary for the distribution of malaria pathogen (*P. falciparum*), and (iv) the suitable niche areas where the ethnobotanical plant species (*L. javanica*) used for its aromatic function is more likely to be found.

Mapping the potential breeding sites for malaria vector in this study will contribute to the increased understanding of outdoor vector habitat distribution and will inform the health professionals about where to look for identifying possible *Anopheles* breeding sites and to concentrate the limited malaria vector control measures. This finding is very crucial particularly that South Africa is also one of the sub-Saharan African countries with the malaria elimination target set to the year 2020. This study has also shown that malaria distribution is highly associated with healthy vegetation with high canopy moisture content. This finding is in agreement with findings from previous studies elsewhere, which highlighted the importance of vegetation and moisture to malaria transmission (Imponvill *et al.*, 2004; Nygren *et al.*, 2014). On the other hand, the distribution of malaria control species commonly used by the local community in rural villages has shown association with slope, elevation, and vegetation greenness. This essentially implies that the likelihood of finding *L. javanica* increase with increase in the above-mentioned variables, essentially implying that this species co-exists with other species in its natural habitat (by virtue of positive correlations between species distribution and vegetation greenness).



However, mapping of malaria using remote sensing is not without challenges. In previous sections (Chapter 2) it was mentioned that challenges in remote sensing for malaria mapping may be associated with data and methods. In this study, the medium resolution Landsat 5 with 30 m spatial resolution and high resolution (10 m) Sentinel-2 were used to map the distribution of malaria and the malaria vector habitats. The Landsat 5 data might not have adequately captured the biophysical and biochemical characteristics of malaria vector habitats due to the fact that 30 m pixel may comprise of many plant species including herbaceous and grass species that can inhibit *Anopheles* oviposition. However, the present study did not focus on the habitat characterization at species level, instead it focused on spectral characteristics at pixel level. Moreover, because of the Landsat spatial resolution it is possible that the pseudo-absences for malaria pathogen may have been drawn from areas where the malaria pathogen was present but not yet identified as such. On the hand, the spectral resampling of Sentinel-2 may have introduced errors that might have affected the predictions, and such classification errors were quantified for cattle hoofprint estimation (RMSE = 0.2 prints/ 100m<sup>2</sup>), LAI retrieval (RMSE = 0.86 m<sup>2</sup> m<sup>-2</sup>), and *L. javanica* mapping (AUC = 0.89).

Despite these challenges, the current study has demonstrated the application of remote sensing for malaria mapping using the optical satellite data. It employed the use of vegetation and other spectral indices with known correlations to malaria vector habitats. In this thesis, it is recommended that future work should focus on the characterization potential malaria vectors habitats using a multi-seasonal approach and very high spatial resolution datasets. In addition, such studies should also look into relating the collected densities of *An. arabiensis* and *An. fenestus* in the malaria affected areas and to relate such information with the LAI measured through ground-based surveys. The availability of current and future very high resolution commercial satellites such as WorldView-2/3, Pleiades provides opportunities for fine resolution malaria vector characterization at household level.



### 7.3 Abridged Summary

#### **Chapter 1:**

Provides the aim, objectives and study design for characterizing environmental variables for mapping malaria distribution in the Vhembe District Municipality by means of remote sensing techniques.

#### **Chapter 2:**

In this chapter, a comprehensive review of the applications of remote sensing and GIS for mapping the environmental variables affecting malaria distribution has been done. It was found that cattle hoofprints, puddles, river edges, low-lying irrigated field, vegetation around homesteads serve as the breeding habitats and questing sites for *An.arabiensis* and *An. fenestus*. This, however, calls for targeted mapping of these environmental conditions as a way to fight malaria transmission in the study area.

#### **Chapter 3:**

In summary, this chapter provides remote sensing methods for mapping of cattle hoofprints in the Vhembe District Municipality. The mapping approach was done through geo-statistics (kriging, co-kriging) and the stepwise multiple linear regression model. It was found that the remote sensing data derived from Sentinel-2 narrow bands are the best predictors of cattle hoof prints, because of their ability to resolve for objects that are associated with cattle hoofprints, e.g. bare soil and wet environments with surrounding vegetation.

#### **Chapter 4:**

In this chapter, the biophysical characteristic of vegetation (including those surrounding villages) in the form of leaf area index (LAI) was retrieved. The retrieval of LAI allows for inference of how much of vegetation leaf surface area contributes to the questing and resting habitats of mosquitoes common in the study area. The results from this study suggest that the broad band vegetation indices that are sensitive to changes in effective LAI were better predictors of remotely-sensed LAI than the narrow band ones. This is an important finding especially considering that the broad bands of Sentinel-2 satellite are configured at a high special resolution of 10 m. Knowing the distribution of LAI at the study area gives the indication of relationship between malaria transmission rates and the outdoor questing/resting areas of mosquitoes.

### **Chapter 5:**

The aim of this chapter was to map the distribution of malaria pathogen *P. falciparum* in the Vhembe District using the Landsat-derived environmental variables. In this chapter, it was observed that the spectral indices such as the soil-adjusted vegetation index (SAVI), and the canopy moisture indices were significantly correlated to the *P. falciparum* occurrence particularly in areas where the pseudo-absences of malaria pathogen were generated at 10 km distance from the known presence locations.

### **Chapter 6:**

In this chapter, we have attempted to map the distribution of ethnobotanically significant plant species (*L. javanica*) across the study area. The occurrence and utilization of this species by local communities contributes to low malaria transmission rates and thus assist in the realization of malaria eradication plans by the World Health Organization (WHO). From the analysis, it was found that the ensemble species distribution model with unequal weighting of remotely-sensed data was more efficient in mapping *L. javanica* than both maxent and logistic regression models.

### **Chapter 7:**

This chapter summarizes findings made in all chapters. From the previous chapters, it was found that both Landsat and Sentinel datasets provide meaning contribution to characterizing the environmental variables that are correlated to malaria transmission. For example, the mapping of cattle hoofprints distribution, LAI retrieval, distribution of *P. falciparum* and mapping of *L. javanica* plant species provide us with insight of how earth observation technology can be applied in epidemiology and to support initiatives meant to curb infectious diseases. This chapter also shows that a lot needs to be done in order to win the battle against malaria in southern Africa, through integration of various technological methods and datasets.

## **7.4 References**

- Adeola A.M., Botai O.J., Olwoch J.M., Rautenbach C.J., Adisa O.M., Taiwo O.J., Kalumba A.M. (2016). Environmental factors and population at risk of malaria in Nkomazi municipality, South Africa. *Tropical Medicine and International Health*, 21(5): 675 – 686.



- Adjorlolo C., Mutanga O., Cho M.A., Ismail R. (2013). Spectral resampling based on user defined inter-band correlation filter: C3 and C4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation*, 21: 535-544.
- Ansari M.A., Vasudevan P., Tandon M., Razdan R.K. (2000). Larvicidal and mosquito repellent action of peppermint (*Mentha piperita*) oil. *Bioresearch Technology*, 71: 267-271.
- Álvarez-Martínez J., Jiménez-Alfaro B., Barquín J., Ondiviela B., Recio M., Silió-Calzada A., Juanes J.A. (2017). Modeling the area of occupancy of habitat types with remote sensing. *Methods in Ecology and Evolution*, 9(3): 580 – 593.
- Dalrymple U., Mappin B., Gething P.W. (2015). Malaria mapping: understanding the global endemicity of *falciparum* and *vivax* malaria. *BMC Medicine*, 13(140): <https://doi.org/10.1186/s12916-015-0372-x>
- Garcia R.A., Lee Z., Hochberg E.J. (2018). Hyperspectral Shallow-Water Remote Sensing with an Enhanced Benthic Classifier. *Remote Sensing*, 10(147).DOI: 10.3390/rs10010147.
- Gwitira I., Murwira A., Masocha M., Zengeya F.M., Shekede M.D., Chirenda J., Tinago W., Mberikunashe J., Masendu R. (2018). GIS-based stratification of malaria risk zones for Zimbabwe. *Geocarto International*. DOI: [doi.org/10.1080/10106049.2018.1478889](https://doi.org/10.1080/10106049.2018.1478889).
- Hirano A., Madden M., Welch R. (2003). Hyperspectral image data for mapping wetland vegetation. *Wetlands*, 23(2): 436 – 448.
- Impoinvil D.E., Kengere J.O., Foster W.A., Iru B.N., Killeen G.F., Githure J.I., Beier J.C., Hassanali A., Knols B.G.J. (2004). Feeding and survival of the malaria vector *Anopheles gambiae* on plants growing in Kenya. *Medical and Veterinary Entomology*, 18: 108-115.
- Kempeneers P., Sedano F., Seebach L., Strobl P., San-Miguel-Ayanz J. (2011). Data Fusion of Different Spatial Resolution Remote Sensing Images Applied to Forest-Type Mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12): 4977 – 4986.
- Khosa E., Kuonza L.R., Kruger P., Maimela E. (2013). Towards the elimination of malaria in South Africa: a review of surveillance data in Mutale Municipality, Limpopo Province, 2005 to 2010. *Malaria Journal*, 12(1):7.
- Kross A., McNairn H., Lapen D., Sunohara M., Champagne C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and



soybean crops. *International Journal of Applied earth Observation and Geoinformation*, 34: 235 – 248.

- Kweka E.J., Mosha F., Lowassa A., Mahande A.M., Kitau J., Matowo J., Mahande M.J., Massenga C.P., Tenu F., Feston E., Lyatuu E.E., Mboya M.A., Mndeme R., Chuwa G., Temu E.A. (2008). Ethnobotanical study of some of mosquito repellent plants in north-eastern Tanzania. *Malaria Journal*, 7(152): <https://doi.org/10.1186/1475-2875-7-152>.
- Lengeler, C. (2004) Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database System Reviews* 2, CD000363.
- Mabogo D.E.N. (1990). The Ethnobotany of the Vhavenda. M.Sc. Thesis. University of Pretoria.
- Malahlela O.E., Adjorlolo C., Olwoch C. (2018). Mapping the spatial distribution of *Lippia javanica* (Burm. f.) Spreng using Sentinel-2 and SRTM-derived topographic data in malaria endemic environment. *Ecological Modelling*, 392: 147 – 158.
- Minakawa N., Sonye G., Mogi M., Githeko A., Yan G. (2002). The Effects of Climatic Factors on the Distribution and Abundance of Malaria Vectors in Kenya. *Journal of Medical Entomology*, 39(6): 833 – 841.
- Mutanga O., Skidmore A.K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19): 3999 – 4014.
- Nygren D., Stoyanov C., Lewold C., Mansson F., Miller J., Kamanga A., Shiff C.J. (2014). Remotely-sensed, nocturnal, dew point correlates with malaria transmission in Southern Province, Zambia: a time-series study. *Malaria Journal*, 13(231). DOI: 10.1186/1475-2875-13-231.
- Okami S., Kohtake N. (2017). Spatiotemporal Modeling for Fine-Scale Maps of Regional Malaria Endemicity and Its Implications for Transitional Complexities in a Routine Surveillance Network in Western Cambodia. *Frontiers in Public Health Health*, 5(262). DOI: 10.3389/fpubh.2017.00262.
- Paaijmans K.P., Thomas M.B. (2011). The influence of mosquito resting behaviour and associated microclimate for malaria risk. *Malaria Journal*, 10(183): DOI: 10.1186/1475-2875-10-183.
- Peng Y., Fan M., Song J., Cui T., Li R. (2018). Assessment of plant species diversity based on hyperspectral indices at a fine scale. *Nature Scientific Reports*, 8:4776. DOI:10.1038/s41598-018-23136-5.

- Peon J., Recondo C., Fernandez S., Calleja J.F., De Miguel E., Carretero L. (2017). Prediction of Topsoil Organic Carbon Using Airborne and Satellite Hyperspectral Imagery. *Remote Sensing*, 9(12): 1211. DOI: 10.3390/rs9121211.
- Pierre-Louis A-M, Qamruddin J, Espinosa I, Challa S. (2018). The Malaria Control Success Story. Chapter 23: 417-432. World Bank.
- Rajan S., Ghosh J., Crawford M.M. (2008). An Active Learning Approach to Hyperspectral Data Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(4). DOI: 10.1109/TGRS.2007.910220.
- Richards S.L., Lord C.C., Pesko K.N., Tabachnick W.J. (2010). Environmental and Biological Factors Influencing *Culex pipiens quinquefasciatus* (Diptera: Culicidae) Vector Competence for West Nile Virus. *American Journal of Tropical Medicine and Hygiene*, 83(1): 126 – 134.
- Seyoum A., Pålsson K., Kung'a S., Kabiru E.W., Lwande W., Killeen W.F., Hassanali A., Knots B.G.J. (2002). Traditional use of mosquito-repellent plants in western Kenya and their evaluation in semi-field experimental huts against *Anopheles gambiae*: ethnobotanical studies and application by thermal expulsion and direct burning. *Journal of Urology*, 96(3): 225 – 231.
- WHO (2008) World Malaria Report 2008, WHO.

### **Curriculum Vitae**

**Oupa Malahlela** was born in Thakgalane, Limpopo province in 1986. He attended Malebala Primary School (1993-1999) and Lephai Secondary School (2000-2004). He received the B.S.c degree in Geography and Environmental Studies from the University of Limpopo, Mankweng, South Africa, in 2009, and the M.Sc Geography at the University of KwaZulu-Natal, Pietermaritzburg, South Africa in 2014. He is currently enrolled for a PhD Geoinformatics at the University of Pretoria, Gauteng South Africa.

In 2010, he joined the then Satellite Applications Center (now South African National Space Agency's Space Operations) at the Council for Scientific and industrial Research, as a GIS/Remote Sensing Intern. In July 2011, he moved to the Natural Resources and Environment division at the CSIR on a Master studentship. He then joined the South African National Space Agency (SANSA) Earth Observation Directorate as a remote sensing scientist, in the Research and Applications Development Unit until April 2019. He is currently employed by the University of Venda as a Lecturer at the Department of Ecology and Resource Management. His current research interests include remote sensing for water inland water mapping, irrigation mapping, water quality assessment, environmental pollution assessment, change detection, remote sensing for health application (water-borne diseases) and invasive species mapping. His research is published in a number of peer-reviewed internal and local journals including the International Journal of Applied Earth Observation and Geoinformation, and the South African Journal of Science. Mr. Malahlela is now registered with the South African Council for Natural Scientific Professions as a Professional Natural Scientist under Geospatial Science category (Reg. No.119345).

---