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

Oupa E. Malahlela,[1,2] Jane M. Olwoch,[1,3] and Clement Adjorlolo[2]

[1]*Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa*

[2]*South African National Space Agency (SANSA), Earth Observation Directorate, Pretoria 0001, South Africa*

[3]*Southern African Science Service Center for Climate Change and Adaptive Land Management (SASSCAL), Windhoek 91100, Namibia*

Abstract:

Objectives: 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 a mean incidence rate of 328.2 per 100 0000 persons/year. This study aimed at evaluating environmental covariates, such as vegetation moisture and vegetation greenness, associated with malaria vector distribution for better predictability towards rapid and efficient disease management and control. **Methods:** The 2005 malaria incidence data combined with Landsat 5 ETM were used in this study. 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 the spatial distribution of malaria in the area. **Results:** A buffer distance of 10 km yielded the highest classification accuracy of 82% at a threshold of 0.9. This model was significant ($\rho < 0.05$) and yielded a deviance (D^2) of 36%. The significantly positive relationship ($\rho < 0.05$) between the soil-adjusted vegetation index (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. **Conclusions:** 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 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

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). Malaria cases are caused by Plasmodium *falciparum* pathogen, which is transmitted by female anopheles mosquitoes to humans as natural intermediate. Malaria is commonly transmitted through the bite of female *Anopheles* mosquitoes on humans. 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 the 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 a zero-malaria country. Such 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, (iii) complete elimination (no transmission) and (iv) prevention of re-introduction of malaria diseases (NDoH, 2010; Maharaj et al., 2012; Mendis et al., 2009). The ultimate goal of these efforts 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 on 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 a 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; Conley et al., 2014). 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 the 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 el., 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 ≤ 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 the authors' knowledge, no study was found in the literature that sought to assesses the efficacy of Landsat derived environmental covariates for predicting malaria when pseudo-absences are derived at various localized buffer distances (<50km).. In the Limpopo province of South Africa, malaria prevalence continues to exert pressures on 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 a mean incidence rate of approximately 328.2 per 100 000 persons/year (Gerritsen et al., 2008). Therefore, this study aimed at evaluating efficacy 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.

2. METHODS

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 (fig. 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 (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 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.

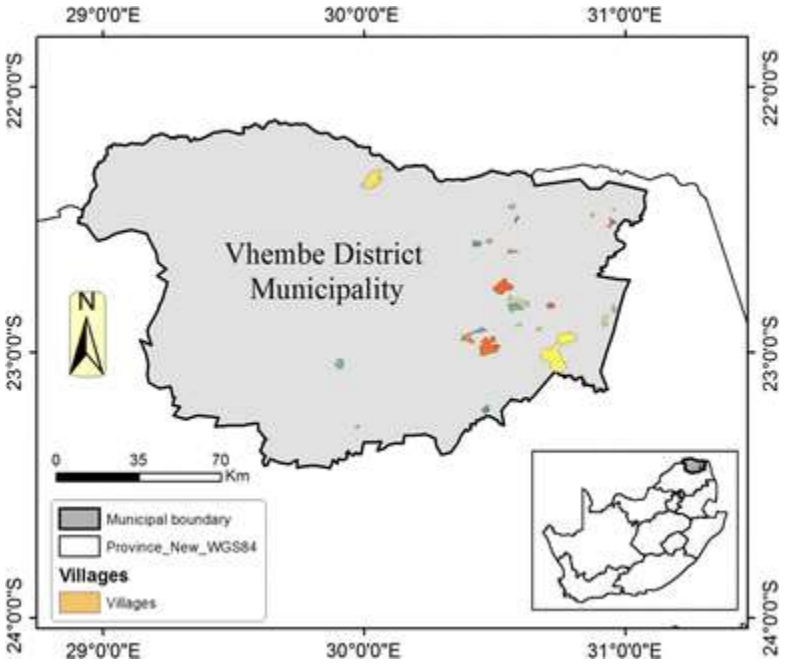

Fig. 1. Study area in the northern part of Limpopo. Each coloured polygon represents individual villages under consideration (n = 28)

2.2. Epidemiological data

This study 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 study was obtained from the faculty of Natural and Agricultural Science Ethical Committee at the University of Pretoria. 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.

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 the 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 µm) 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.

2.4. Data pre-processing

Figure 2 shows the workflow adopted in this study. 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 contains 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 to 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 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).

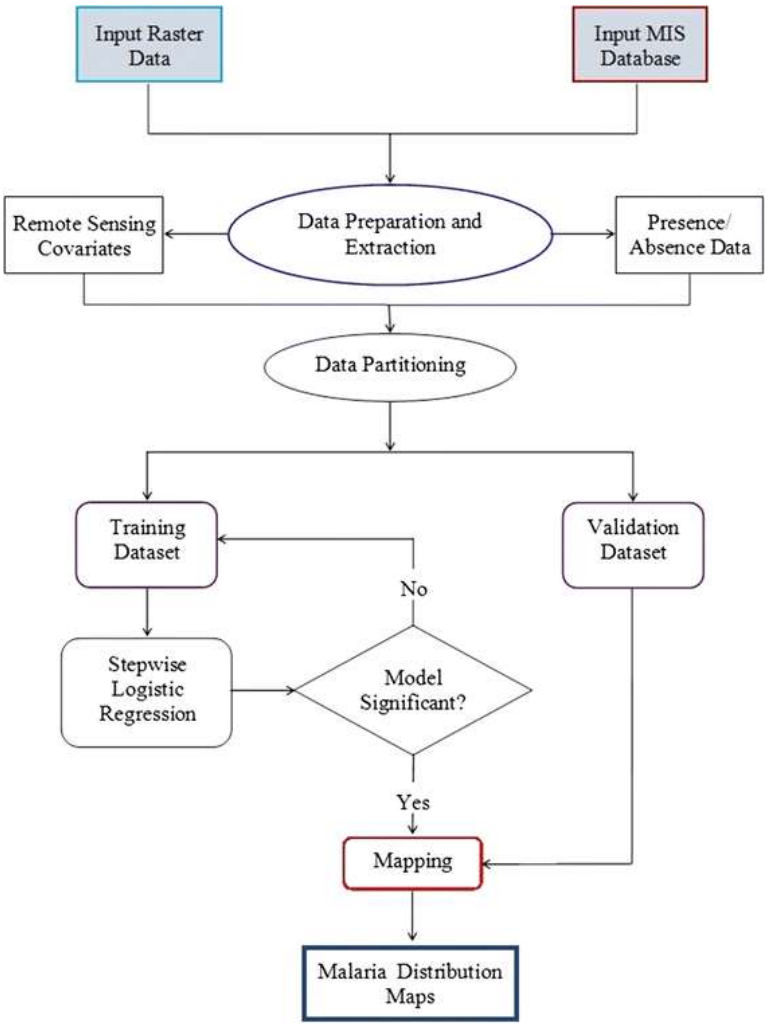

Fig. 2. Overview of epidemiological and remote sensing datasets and methods used for the study

2.5. Data processing and analysis

The pseudo-absences were generated for five buffer distances from the known presence locations. To test for the efficacy 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 (fig. 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.

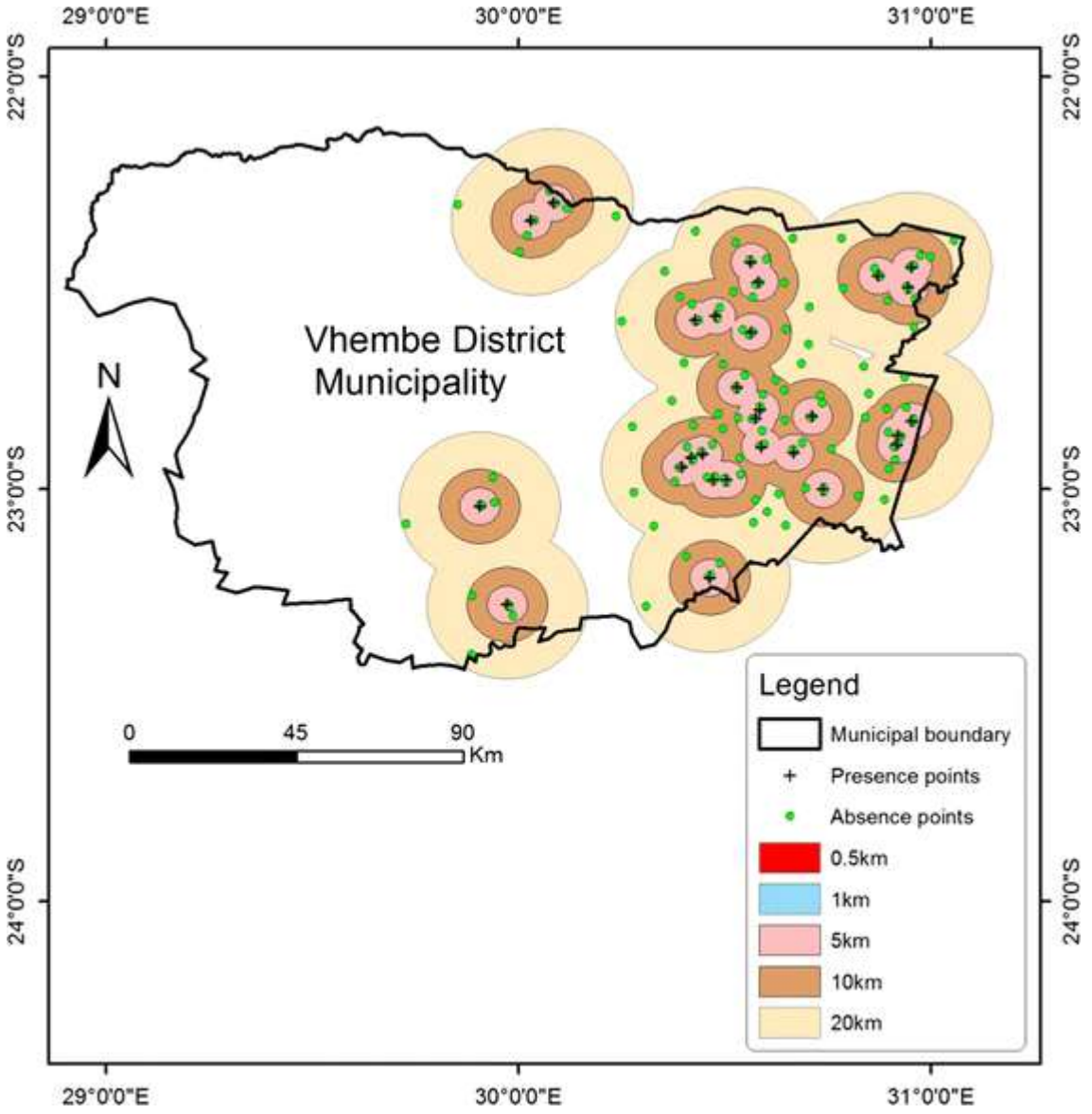

Fig. 3. Representation of different buffer distances across 28 villages in Vhembe District Municipality. Buffer distances were ranging from 0.5 to 20 km

2.6. Remote sensing 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 ($MNDWI_1$ and $MNDWI_2$), 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 a 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. 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 2. These indices were computed by considering the reflectances in the shortwave infrared spectral region (1.55-2.35 μm) to malaria mapping, which may have the potential confounding effect to $MNDWI_1$. The shortwave infrared (SWIR) reflectance generally decreases as water content in the leaves increases at 1-3μm (Gao, 1996; Hunt and Rock 1989) and in this study 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 a similar spatial resolution as Landsat TM of the study area.

Table 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)	$\mathrm{NDVI} = \frac{(\rho_{NIR} - \rho_{Red})}{(\rho_{NIR} + \rho_{Red})}$	Jackson et al. (1983)
Modified Normalized Difference Water Index ($MNDWI_1$)	$\mathrm{MNDWI}_1 = \frac{(\rho_{NIR} - \rho_{SWIR})}{(\rho_{NIR} + \rho_{SWIR})}$	Ceccato et al. (2001)
Modified Normalized Difference Water Index ($MNDWI_2$)	$\mathrm{MNDWI}_2 = \frac{(\rho_{Green} - \rho_{MIR})}{(\rho_{Green} + \rho_{MIR})}$	Xu (2006)
Moisture index (a_1)	$a_1 = \left(\frac{\rho_{SWIR} + \rho_{MIR}}{2}\right)$	*In this study*
Moisture index (a_2)	$a_2 = \rho_{NIR} - \rho_{SWIR}$	*In this study*
Soil adjusted vegetation index	$\mathrm{SAVI} = \frac{(\rho_{NIR} - \rho_{Red})}{(\rho_{NIR} + \rho_{Red} + L)}(1 + L)$	Huete (1988)
Green index	$GI = \frac{\rho_{NIR}}{\rho_{Green}}$	Gitelson et al. (2003)
Quasi-Yellowness Index	p-$YI = \frac{\rho_{Yellow} - (2 * \rho_{Green}) + \rho_{Red}}{{\rho_{Blue}}^2}$	Adams et al. (1999)

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 Core Development 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 in R, 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 (1):

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

where y_i is the probability of malaria distribution (1 or 0), x_i is the environmental covariate at the *j*th location, β_i is the coefficient of x_n, β_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 Aikaike'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 (2):

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

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. Because logistic regression is a form of generalized linear model with binary output the model D^2 is used. The D^2 is derived from the null deviance, which measures the variability of dataset, compared to the residual deviance, which measures variability of the residuals after fitting the model. These deviances can be interpreted much like the total and residual sum of squares in a linear model to estimate the goodness of fit (Rossiter and Loza, 2016). 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).

3. RESULTS

The results of the study are presented in table 2. The table showed that SAVI was the most significant remotely-derived covariate in predicting the distribution of malaria in VDM, Limpopo.

Table 2: Results of the logistic regression along 5 buffer distances from the known *P. falciparum* locations.

Buffer distance (km)	Covariates	Estimates	ρ value (> | z |)	Significance level
0.5	β_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	β_0	0.629	0.755	
	SAVI	14.882	0.041	*
	$NDWI_1$	45.984	0.078	
	a_2	87.529	0.094	
	p-YI	-7.241	0.013	*
5.0	β_0	5.341	0.001	***
	SAVI	13.659	0.012	*
	Aspect	0.007	0.061	
10.0	β_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	
	$NDWI_1$	89.790	0.049	*
	$NDWI_2$	-36.030	0.016	*
20.0	β_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) (ρ <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 (Fig. 4), while at 5 km low classification accuracy (54%) was found at the threshold value of ρ = 1.0.

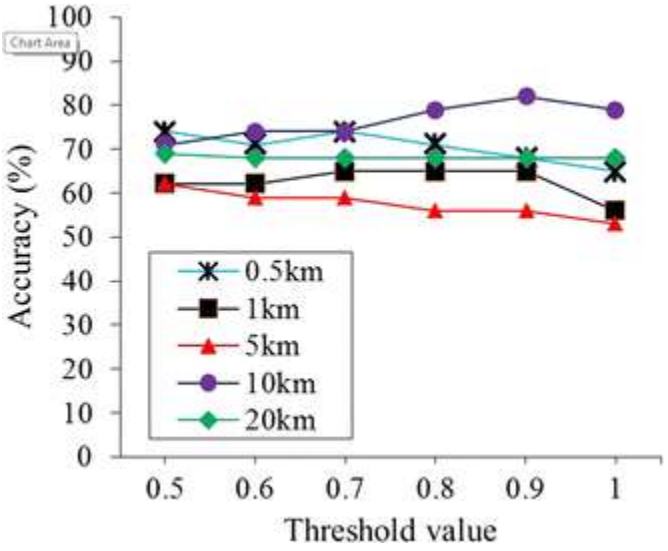

Fig. 4. Logistic regression performance across threshold values of 0.5–1.0 as applied on buffer distances selected for the study area using validation dataset (n = 34)

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 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.

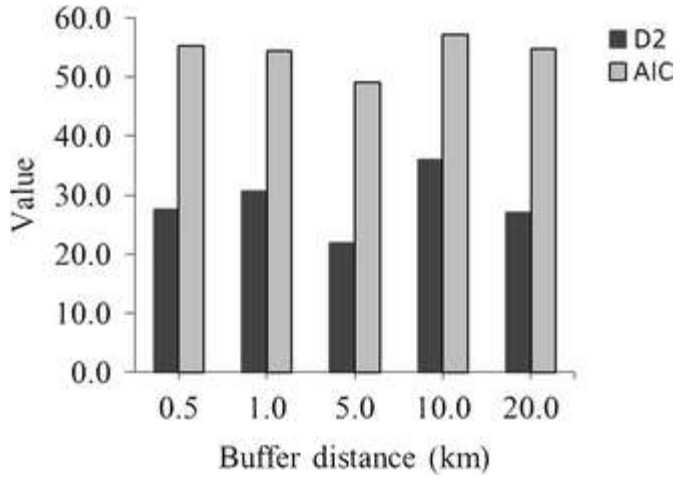

Fig. 5. D^2 and AIC of logistic regression applied on buffer distances in VDM

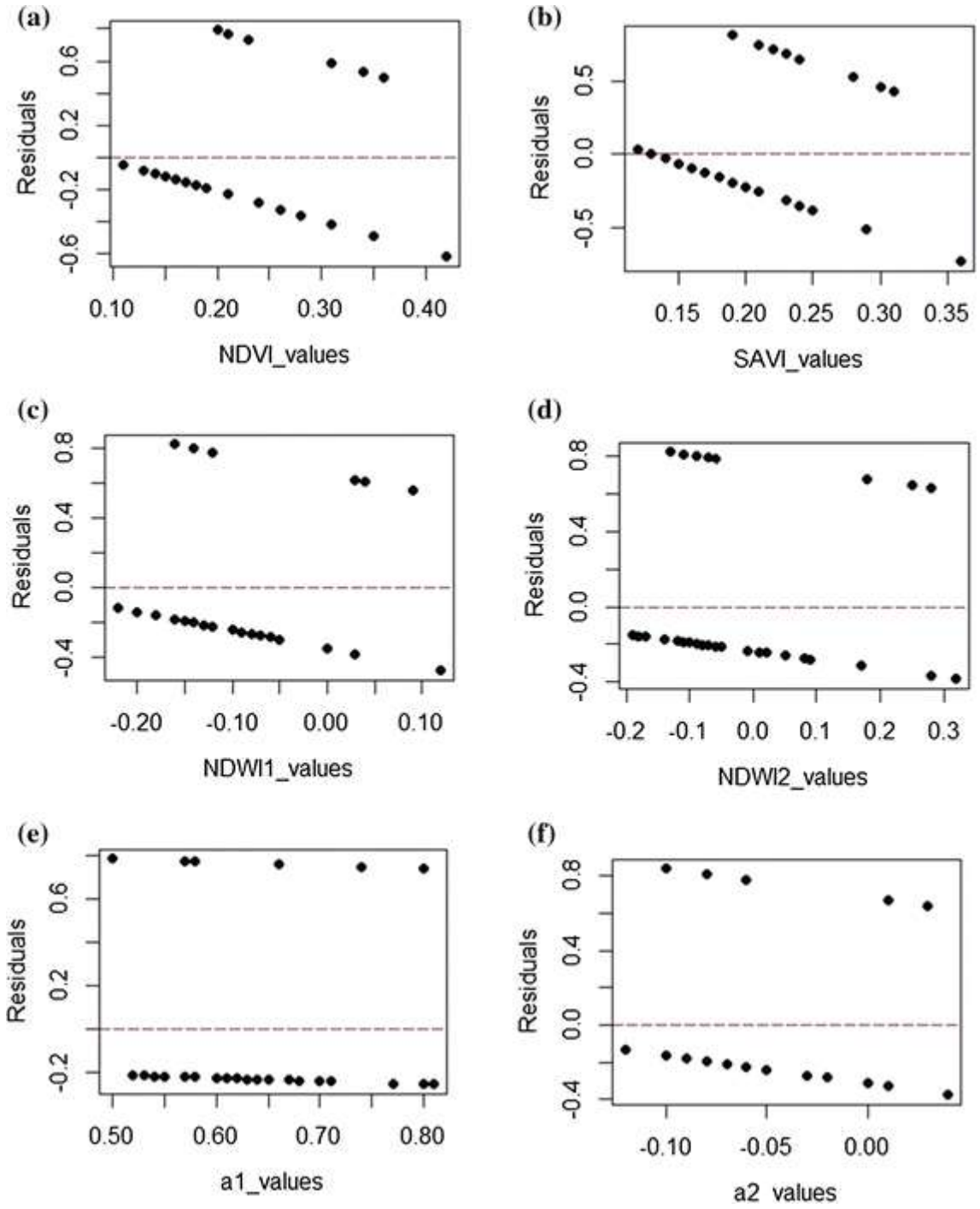

Fig. 6. Residual plots of environmental covariates of the significant SLR model

The F-test performed between the measured and predicted presence/absence data yielded a statistical ρ value of 0.081, which is greater than the $\rho = 0.05$. 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 ρ=0.033. The variable residual plots (Fig. 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 with lower residuals 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.

3.1. Predictive maps

The results of the predicted probability maps are shown in figures 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 pattern in malaria distribution (fig. 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 (fig. 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 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.

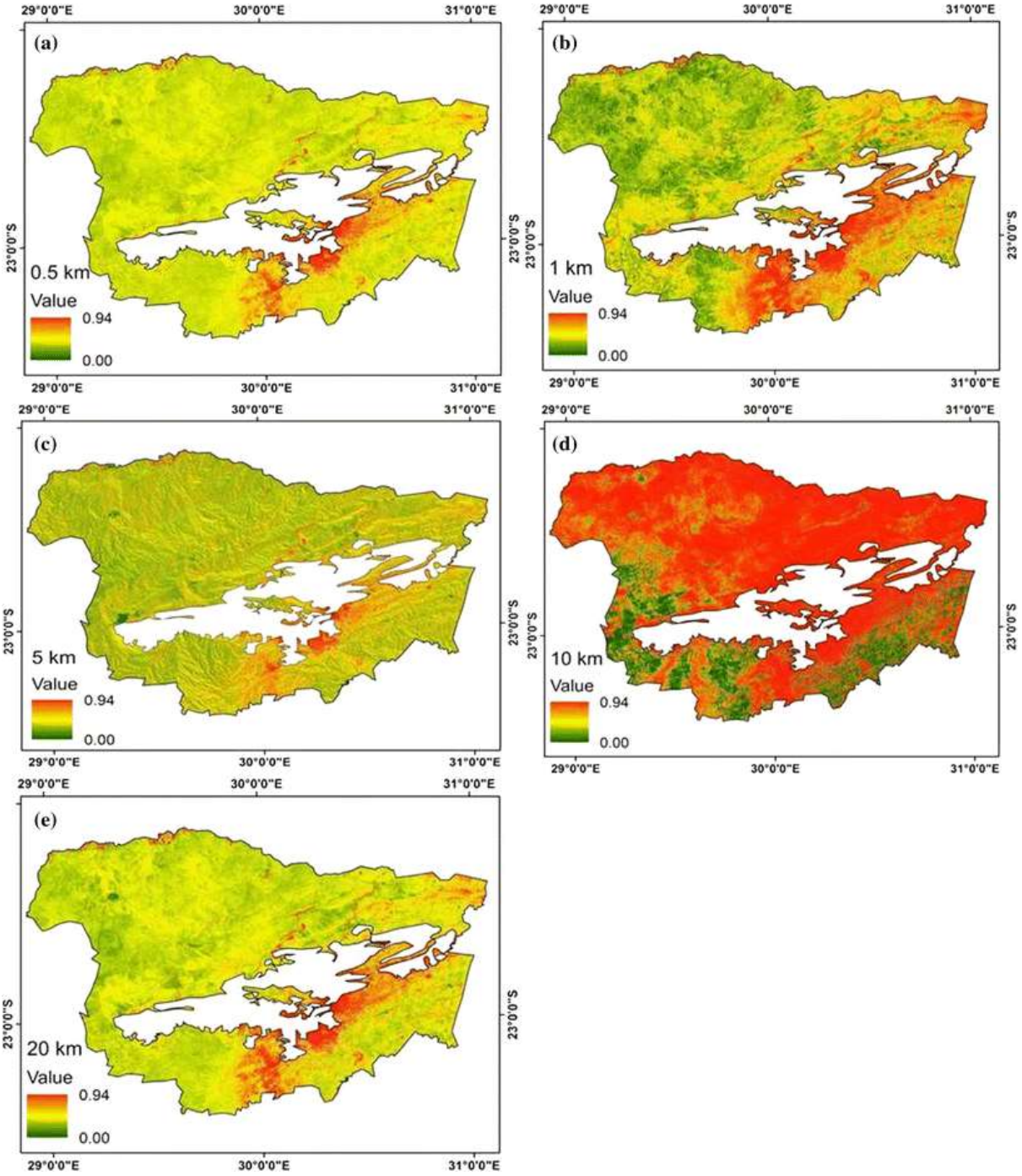

Fig. 7. Predicted potential geographic distribution of malaria produced through logistic regression and Landsat-derived environmental covariates. Maps produced at buffer distances of 0.5 km (**a**), 1 km (**b**), 5 km (**c**), 10 km (**d**) and 20 km (**e**)

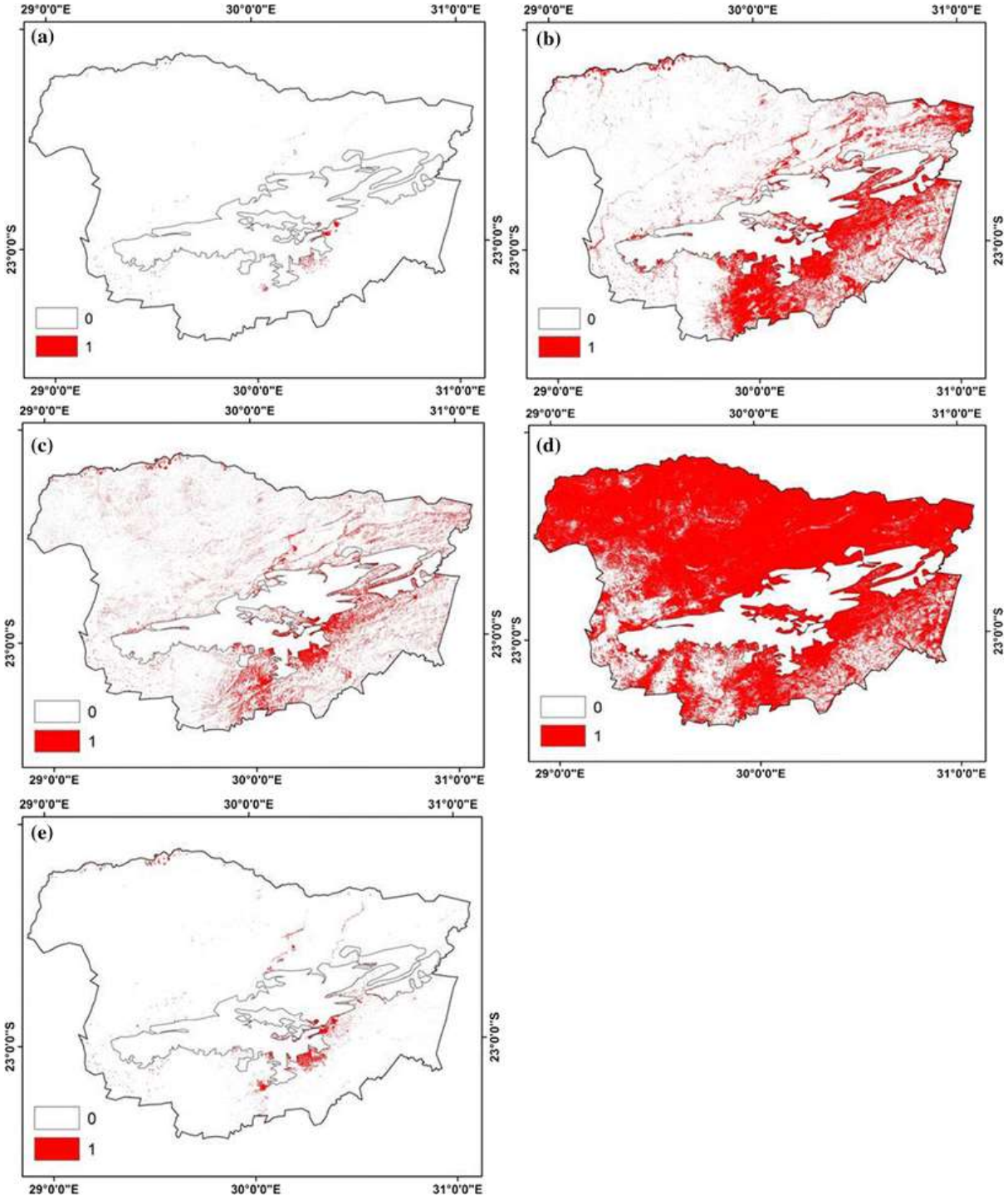

Fig. 8. Predicted spatial distribution of malaria in VDM at a probability threshold of ≥ 0.7. Red colour depicts areas of predicted *P. falciparum* presence, while white represents predicted *P. falciparum* absence and Soutpansberg Mountain mask (Color figure online)

4. DISCUSSIONS

The aim of this study was to assess the feasibility of Landsat-derived environmental covariates for predicting malaria distribution in the rural landscape of Vhembe District Municipality in South

Africa, across different buffer distances. In this study, the results have shown that intermediate buffer distance (10 km) yielded the highest classification accuracy (82%) at the threshold of 0.9, with a D^2 of 0.36. 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, ρ = 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.

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 study, 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 *Anopheline* 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 a South African semi-arid environment that assessed the correlation of remote sensing covariates to *P. falciparum* distribution at a 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, this study has shown that the NDVI correlation is environment dependant and therefore in semi-arid environments, SAVI which takes into account effect of soil background, performs higher than NDVI. The 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 significantly 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 study 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 $MNDW_1$ 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 $MNDWI_1$ 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 $MNDWI_1$ 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 $MNDWI_1$, 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.

The inability of the significant remotely-sensed covariates to explain high probability (D^2 <50%) of occurrence of malaria distribution is likely due to the complexity of other unexplained variables relating to *Anopheles* breeding sites and human factors that were not measured in this study (Protopopoff et al., 2009). There is also an increasing recognition that the dynamics of infectious malaria transmission is as a result of the complex interplay between human, animal and environmental conditions (Ganser and Wisely, 2013). Assessing the interlink of variables from all key contributory factors in malaria transmission could effectively improve the model accuracy (Dlamini et al., 2015). On the other hand, the significance of SAVI, NDVI, $NDWI_1$ and $NDWI_2$ in the final predictive model is an indication that remote sensing variables are strongly correlated to the malaria occurrence although the variation explained is lower. In addition, a multi-temporal analysis of malaria transmission in VDM by use of remote sensing could potentially shed the light on effectiveness of this technology at district level. This is primarily due to the seasonality of malaria in South Africa and the neighbouring countries especially in Zimbabwe (Mabaso et al., 2005).

The pattern of malaria distribution is highlighted by the probability maps in figure 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 (fig. 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 fig. 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 on 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 a 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 southwestern part of Soutpansberg Mountain.

Although this study 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 ρ 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 study. Malaria transmission in 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 the availability of 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. CONCLUSIONS

In conclusion, findings from this study show 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 an 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 study 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.

Acknowledgements

Authors would like to thank QGIS Development Team for the Quantum GIS software used in this study. We would like to also extent our gratitude to the R Development Team for making R software available for data analysis. Our gratitude to the USGS for free Landsat dataset used in the study. This work is funded by the South African National Space Agency under the Human Capital Development. Authors would also like to thank two anonymous reviewers who have helped improve the quality of this manuscript.

Authors' Contribution

OM, JO, CA conceptualized the research. JO and CA participated in data analysis. OM drafted the manuscript. JO and CA supervised the entire work. All authors read and approved final manuscript.

Conflict of interest

Authors declare that they have no conflict of interest.

Acronyms

AIC:	Aikaike's Information Criterion
ASTER:	Advanced Spaceborne Thermal Emission and Reflection Radiometer
D^2:	Deviance
DEM:	Digital elevation model
GARP:	Genetic Algorithm for Rule-Set Prediction
GI:	Greenness index
MIS:	Malaria Information System
MNDWI:	Modified normalized difference water index
NDVI:	Normalized difference vegetation index
p-YI:	Quasi-Yellowness Index
SAVI:	Soil-adjusted vegetation index
SDM:	Species Distribution Modelling
SLR:	Stepwise logistic regression
TM:	Thematic Mapper
VDM:	Vhembe District Municipality

References

1. Adams J. (2010). Vegetation–climate interaction. ''How plants make the global environment'', 2nd Edn. Springer, New York. ISBN:978-3-642-00880-1.
2. 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
3. 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.
4. Adimi F., Soebiyanto R.P., Safi N., Kiang R. (2010). Towards malaria risk prediction in Afghanistan using remote sensing. *Malaria Journal* 9:125.
5. 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.
6. 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.
7. Araujo S. L., Santos J.R., Shimabukuro Y.E. (2000). Relationship between SAVI and biomass data of forest and savannah contact zone in the Brazillian Amozonia. *International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII*, Part B7. Amsterdam.
8. 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.
9. Baldwin R.A. (2009). Use of maximum entropy modelling in wildlife research. *Entropy* 11:854–866.
10. 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.
11. 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.
12. 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.

13. 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.
14. 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.
15. 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.
16. 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.
17. Chefaoui R.M., Lobo J.M. (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological ModelingModelling* 210: 478 – 486.
18. 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.
19. Collett D. (1991). Modeling binary data. Chapman and Hall, London.
20. 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.
21. 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.
22. 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.
23. ENVI (2009). Atmospheric correction module: QUAC and FLAASH user's guide. Version 4.7. 20AC47DOC.
24. 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.
25. Exelis Visual Information Solutions. (2016). Environment for Visualizing Images. Boulder, CO: Exelis Visual Information Solutions.
26. Ganser C., Wisely S.M. (2013). Patterns of spatio-temporal distribution, abundance, and diversity in a mosquito community from the eastern Smoky Hills of Kansas. *Journal of Vector Ecology* 38(2): 229 – 236.
27. 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.
28. 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.
29. 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.
30. Huete A.R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25(3):295–309.
31. 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.
32. 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.
33. 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.
34. 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.*

35. 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.
36. 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.
37. 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.
38. Komen K. (2017). Could Malaria Control Programmes be timed to coincide with onset of rainfall? *EcoHealth* 14(2): 259 – 271.
39. 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 smiple econometric approach. *EcoHealth* 12(1): 131 – 143.
40. Mabaso M.L.H., Craig M., Vounatsou P., Smith T. (2005). Towards empirical description of malaria seasonality in southern Africa: the example if Zimbabwe. *Tropical Medicine and International Health* 10(9): 909 – 918.
41. 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.
42. Machault V., Vignolles C., Borchi F., Vounatsou P., Pages F., Briolant S., lacaux J., Rogier C. (2011). The use of remotely sensed environmental data in the study of malaria. *Geospatial Health* 5: 151 – 168.
43. Machault V., Vignolles C., Pagès F., Gadiaga L., Gaye A., Sokhna C., Trape J., Lacaux 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.
44. 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.
45. Moyes C.L., Shearer F.M., Huang Z., Wiebe A., Gibson H.S., Mohd-Azlan J., Brodie J.F., Malaivitnond 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.
46. 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.
47. National Department of Health. (2011). *Malaria Elimination Strategy 2011 – 2018 in South Africa.* South Africa: National Department of Health (NDoH).
48. National Institute for Communicable Diseases (NICD)-National Health Laboratory Services (NHLS) (2017). Malaria in South Africa: An Update. 16(5):1
49. 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.
50. 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.
51. Phillips S.J., Anderson R.P., Schapire R.E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling* 190: 231–259.
52. 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.

53. Protopopoff N., van Bortel W., Speybroeck N., van Geertruyden J., Baza D., D'Alessandro U., Coosemans M. (2009). Ranking Malaria Risk Factors to Guide Malaria Control Efforts in African Highlands. *PLoS One* : doi.org/10.1371/journal.pone.0008022
54. QGIS Development Team. (2016). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.
55. 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.
56. 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.
57. Ripley B.D. (2003). Selecting amongst large classes of models. Lecture, URL http://www.stats.ox.ac.uk/~ripley/Nelder80.pdf
58. Rossiter D.G., Loza A. (2016). Technical Note: Analysing land cover change with logistic regression in R. 1 – 67. http://www.css.cornell.edu/faculty/dgr2/teach/R/R_lcc.pdf . Accessed on the 30/10/2017.
59. 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 bionomic précis. *Parasites and Vectors* 3: 1 – 26.
60. Statistics South Africa. (2016). Population census 2016. Vhembe District Municipality. http://cs2016.statssa.gov.za/?page_id=270
61. Tonnang H.E.Z., Kangalawe R.Y.M., Yanda P.Z. (2010). PreictingPredicting and mapping malaria under climate change scenarios: the potential distribution of malaria vectors in Africa. *Malaria Journal* 9(111): 1 – 10.
62. Tucker C.J. (1980). Remote sensing of leaf water content in the near infrared. *Remote Sensing of Environment* 10(1): 23 – 32.
63. 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.
64. World Health Organization (WHO) (2016). World malaria report 2016. Geneva: World Licence:CC BY-NC-SA 3.0 IGO.
65. 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.
66. 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.
67. 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.