Remote sensing of environmental risk factors for malaria in different geographic contexts

Andrea McMahon
University of Oklahoma
Abere Mihretie
Health, Development, and Anti-Malaria Association
Adem Agmas Ahmed
Malaria Control and Elimination Partnership in Africa
Mastewal Lake
Public Health Institute
Worku Awoke
Bahir Dar University
Michael Wimberly (mcwimberly@ou.edu)
University of Oklahoma https://orcid.org/0000-0003-1549-3891

Research Article

Keywords: NDVI, NDWI, Millennium Development Goals (MDGs), wetlands, seasonally flooded waterbodies

Posted Date: March 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-280021/v1

Abstract

Background

Despite global intervention efforts, malaria remains a major public health concern in many parts of the world. Understanding geographic variation in malaria patterns and their environmental determinants can support targeting of malaria control and development of elimination strategies. We used multiple sources of remotely-sensed environmental data to analyze the influences of environmental risk factors on malaria cases caused by *Plasmodium falciparum* and *Plasmodium vivax* in two geographic settings in Ethiopia.

Results

We found considerable spatial variation in malaria proportion and the distribution of malaria hot spots. Spectral indices were related to land cover greenness (NDVI) and moisture (NDWI) showed an association between malaria and dry landscapes. Climatic factors, including precipitation and land surface temperature, had positive associations with malaria occurrence. Settlement structure also played an important role, with opposing relationships between settlement density and malaria for the two study areas. Variables related to land surface water, such as irrigated agriculture, wetlands, seasonally flooded waterbodies and height above nearest drainage did not influence malaria proportion.

Conclusion

We found different relationships between malaria and environmental conditions in two geographically distinctive areas. These results emphasize that studies of malaria-environmental relationships and predictive models of malaria occurence should be context specific to account for such differences.

Introduction

According to the United Nations Millennium Development Goals (MDGs), combatting diseases, including mosquito borne diseases such as malaria, is a high priority. In particular, malaria is the focus of continued efforts toward control and elimination [1, 2]. There has been significant progress in reducing the burden of malaria, but it remains a major public health concern with 228 million malaria cases and with 405 000 malaria deaths in 2018 [3]. The goal is to reduce these numbers by enabling access to prevention, diagnostic testing and treatment for all people[4]. However, resources to achieve these goals are limited [5]. It is essential to use available resources efficiently by spatially targeting prevention, control, and elimination efforts. Therefore, knowing what drives spatial and temporal patterns of disease risk is crucial when responding to disease outbreaks, and science can support public health efforts by identifying areas with high risk of disease transmission [6–8]. In this study we investigate how satellite remote sensing data can be used to identify environmental risk factors for malaria at local scales that are relevant for targeting malaria interventions.

Mosquito borne diseases, such as malaria, are highly sensitive to environmental conditions. Mosquito life history traits like longevity, fecundity, and biting rates are highly influenced by temperature [9] and humidity [10]. The rate of pathogen development inside the mosquito (extrinsic incubation period) and transmission probabilities between human and mosquito are also influenced by temperature [9]. Larval habitats, and therefore mosquito abundance, are influenced by land cover [11], hydrological setting [12] and water management for irrigation [13]. In African cities, the risk of contracting malaria is associated with dense vegetation cover and swampy areas, while build-up urban areas have found to be associated with lower malaria risk [14]. Monitoring and mapping these meterological and landscape variables can provide information about the times and locations where disease transmission risk is highest.

The use of geospatial environmental datasets to study the risk of diseases, including those transmitted by mosquitoes, has greatly expanded in recent decades [15, 16]. Satellite imagery enable the monitoring of environmental conditions consistently and continuously over large areas. The large number of available sensors allows us to measure a wide range of environmental factors that influence disease transmission, including meterological factors such as temperature, humidity, and precipitation as well as landscape features related to land use, vegetation, surface water, and terrain. This makes remote sensing a useful tool for studying the effect of environmental conditions on mosquito borne diseases in general [17] and malaria in particular [18–21]. In Ethiopia, such studies have established a relationship between malaria risk and remotely-sensed environmental factors, such as land surface temperature [19, 22], precipitation [22, 23], greenness and moisture indices [19], wetland abundance [24], and distance to water bodies [22].

However, the interaction between climatic conditions and mosquito borne disease risk can be heterogeneous across a landscape because land cover and physiography can mediate how climate and malaria risk are related [25]. For example, rainfall was found to be a very strong predictor of mosquito larval abundances in an area with high malaria transmission in Tanzania, but only after taking the geomorphological setting into account and stratifying by water body type, [12]. Precipitation effects on malaria were also medicated by landscape physiography in the Brazilian Amazon, where there was a positive relationship between malaria cases and precipitation in the upland regions and a negative relationship in wetland regions [26]. Similar effects has also been observed in the Amhara region of Ethiopia where malaria cases were influenced more by temperature in wetter locations and influenced more by moisture variables in drier locations [27]. This evidence of geographic variation in the environmental drivers of malaria was used as the basis for developing a genetic algorithm to identify groups of woredas in the Amhara region where malaria outbreaks respond similarly to climate fluctuations [20].

To be most relevant for public health, geographic studies should be conducted at the administrative level that is used for decision making and intervention planning. However, in many cases data are not collected and stored at the level that is most useful for decision makers [28]. In Ethiopia, spatial analyes of malaria often use case data that are aggregated by district (*woreda*) [29, 30], including several previous studies relating malaria cases to environmental risk factors [20, 31]. However, relationships

between environmental factors and diseases can vary at different geographic aggregation levels [28]. This phenomenon is commonly referred to as the Modifiable Area Unit Problem (MAUP) and is of particular concern in a geographically heterogeneous country like Ethiopia. Woredas range from 0.4 km^2 – 13,500 km^2 in size with an average of 1,300 km^2. Their populations ranges from 9,000–586,000 for all woredas in Ethiopia, with a median population of 100,100. Within a woreda, there can be considerable differences in climate, topography, and population density. As established earlier, these environmental factors can influence mosquito populations, pathogen development, as well as malaria transmission to humans. Smaller-scale studies are often based on surveys of households [32]. From such studies, we know that malaria prevalence varies between households within woredas [33, 34], as does the distribution of the two pathogens *P. falciparum* and *P. vivax* [35]. However, household-based studies are usually performed on a limited sample of households within one or a few villages.

There are fewer malaria studies based on sub-district administrative units, hereafter referred to as *kebeles.* Kebeles are the smallest administrative unit in Ethiopia with areas of kebeles for the entire country ranging from 0.1 km^2 – 9,500 km^2 with an average of 50 km^2 with a population of typically a few thousand inhabitants. Studying malaria at this aggregation level allows for the study of larger areas than household-survey studies, while capturing more spatial variation in malaria and the physical environment than woreda-level studies. Previous kebele-level studies have shown that malaria prevalence varies between kebeles within the same woreda in the Amhara [36] and Oromia [34] regions of Ethiopia. However, to our knowledge, no other research has analyzed the environmental factors associated with kebele-level patterns of malaria.

The knowledge that 1) malaria-environment associations can vary considerably between woredas and 2) there is considerable variation in malaria transmission within woredas highlights the need to understand the environmental determinants of malariaat a kebele-level. Our objectives are therefore to characterize the spatial patterns of malaria cases at sub-woreda scales and identify identify the environmental risk factors that are associated with these patterns. We do so by leveraging multiple sources of satellite imagery to estimate a variety of environmental factors related to climate variation, land use and land cover, water bodies, and human settlements.

Methods

Study area

Our study area covers four woredas in the Amhara region of Ethiopia (Figure 1). The Amhara region is located in northwestern Ethiopian and is geographically very diverse. Elevation ranges from 500 m to 4500 m. Climate is very seasonal with a pronounced rainy season and dry season. The four woredas are located in two geographically separated locations. The first study area, encompassing Mecha and Bahir Dar Zuryia, is located south of Lake Tana and the city of Bahir Dar in central Amhara. In these woredas, elevations range from 1500 m to 3200 m, mean daily temperatures range from 16 °C to 26 °C, and annual precipitation is between 1000 mm and 2000 mm. The northern section is relatively flat with large

expanses of mixed agriculture and extensive wetlands, transitioning to the foothills of the Choke Mountains in the southern section. This study area contains a substantial area of irrigated agriculture within the Koga irrigation development project. The Koga project encompasses two dams with respective water detention ponds and 11 night water storage reservoirs. It irrigates 7004 ha of farmland via a system of canals [37].

The second study area, encompassing Aneded and Awabel, is located on the southern edge of the Amhara region between the Choke Mountains to the north and the Blue Nile River to the south. In these woredas, elevations range from 1000 m to 3600 m, mean daily temperatures range from 15 °C to 28 °C, and annual precipitation is between 900 mm and 1300 mm. This northern section has gradually sloping terrain and is covered by a mosaic of croplands and wetlands, while the southern section drops steeply into the Blue Nile Gorge and has some croplands with large areas of bare soil and sparse vegetation. There is no large-scale irrigation scheme in this area.

The Amhara region has unstable malaria transmission that results in sporadic localized and regional outbreaks [38]. Outbreaks occur primarily between September and December. The malaria parasites in this region are *Plasmodium falciparum* and *Plasmodium vivax* [35] and the primary vector species in this region is the mosquito *Anopheles arabiensis* [39]. The Ethiopian government has committed to move towards nationwide malaria elimination by 2030 [40]. Due to successes of malaria control efforts, malaria incidence and mortality rates have been declining in the Amhara region [29] and in other parts of Ethiopia [3]. In the Amhara region, confirmed malaria cases are usually reported from health facilities and then aggregated by woredas under a new malaria elimination pilot project, malaria data from seven woredas were aggregated at kebele level. Each kebele has at least one health post that serves up to 5,000 people. Larger kebeles have health centers that serve 20,000 people. Our study area includes a total of 122 kebeles from the four study woredas: Mecha, Bahir Dar Zuryia, Aneded and Awabel.

Malaria data

Malaria case data were collected at local health posts and health centers, summarized by week for each kebele, and reported to the Amhara Regional Health Bureau. The data included malaria cases of patients who sought care at a health post or health center and were confirmed by rapid diagnostic test or blood film screening. Weekly summaries included information on the malaria-causing pathogen (total *P. falciparum* cases, *P. vivax* cases, and mixed infections), age of the patients (above or below 5 years of age), the number of malaria patients with a travel history, and the number of total outpatients seeking care during a given week. Since no recent population data were available at the kebele level, we calculated the proportion of outpatients diagnosed with malaria (hereafter referred to as malaria proportion). This ratio is considered a reliable indicator of malaria burden because it controls for temporal variation in health facility attendance and can be calculated in situations where accurate population data are not available [41,42]. The case data ranged from September 2013 to July 2018. However, we only used data from 2014 – 2017, as these were available for the entire year. Out of 472

reporting health posts or health centers, we did not have reliable location data for six. Together, these six reported only 23 malaria cases for the entire time frame and their exclusion was not expected to influence the results.

Epidemiological data were summarized for all health posts within a kebele and for each year, to produce a kebele-wide annual tally of total malaria cases. We summarized total malaria cases, as well as for *P. falciparum* + mixed infections combined and *P. vivax* individually. To detect statistically significant spatial clusters in malaria occurrence, we performed a scan statistic using SaTScan software version 9.6 [43]. We ran a retrospective purely spatial discrete Poisson model with total outpatients as the population at risk. The scan statistic was performed with an elliptical window for each year and each malaria pathogen group (total malaria cases, *P. falciparum +* mixed cases, and *P. vivax* cases) separately. After determining the spatial clusters for each year, we then identified stable hot spots (areas with recurring clusters in three or four years), unstable hot spots (clusters in one or two years), and areas that were never identified as clusters.

Environmental variables

Table 1: List of static and dynamic variables used to predict malaria proportion.

Type	Description	Source	Name	Units
Dynamic variables	Daytime temperature	Terra MODIS	LST	°C
	Annual Rainfall	IMERG	PREC	mm
	Normalized Difference Vegetation Index	MODIS NBAR	NDVI	index
	Normalized Difference Moisture Index	MODIS NBAR	NDMI	index
Static variables	Settlement mean density	PlanetScope	SETME	index
	Settlement max density	PlanetScope	SETMX	index
	Area below 2 m above nearest drainage	DEM, Stream network	HAND	%
	Wetland cover	Midekisa et al. (2014)	WETL	%
	Woody vegetation cover	Midekisa et al. (2014)	WOODY	%
	Cropland cover	Midekisa et al. (2014)	CROP	%
	Open water cover	Midekisa et al. (2014)	WATER	%
	Sparse vegetation cover	Midekisa et al. (2014)	SPVEG	%
	Irriagtion cover	Digitized from Google Earth	IRRI	%
	Distance to Seasonal Waterbodies	Landsat OLI	DISTSW	m

Dynamic variables are summarized for each year for the dry season (_dr), rainy season (_rn), and transition season (_tr).

Two types of explanatory environmental variables were used to investigate malaria case patterns (Table 1). *Dynamic variables* included environmental conditions that were expected to vary between and within years, such as land surface temperature and remotely-sensed greenness and moisture indices. Temperature data were derived from the MODIS Terra Land Surface Temperature and Emissivity Product (MOD11A2) [44]. These data have a spatial resolution of 1km and are provided as 8-day composites. We used daytime observations to reduce the problem of missing data from nighttime clouds. Data were filtered using the quality assurance flags to only include observations with an average LST error of below 2 °K. Temperature values were then converted to °C.

Spectral indices measuring vegetation greenness and surface moisture were derived from daily 500 m spatial resolution MODIS Nadir BRDF-Adjusted Reflectance data (MCD43A4) [45]. We applied a data quality filter using the BRDF/Albedo quality product (MCD43A2) and only included observations that were

flagged as land and were “good” or “best” quality. We then calculated the Normalized Difference Vegetation Index (NDVI) [46], as well as a Normalized Difference Moisture Index (NDMI) [47].

We imputed missing values and replaced outliers from imperfect cloud-screening using a robust linear regression model from the R MASS library [48]. We fitted a robust linear regression model on our temporal data for NDVI, NDMI, and LST using cyclical splines and estimated outliers with a z-score above 3 or below -3. We then removed observations that were identified as outliers and replaced them, as well as missing values, with predicted values from the robust linear regression model.

Precipitation data was derived from the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement product (IMERG) [49]. The IMERG product has a spatial resolution of 0.1° and a temporal resolution of 30 minutes. Thee dynamic variables for land surface temperature, spectral indices, and precipitation were summarized for each kebele and for each season of the year: dry season (January – April), rainy season (May – August), and the transition season (September – December).

To identify potential mosquito breeding habitats, we mapped seasonally flooded areas for each year using 30 m Landsat 8 Operational Land Imager (OLI) imagery for the dry season and the rainy season of each year. We removed pixels that were flagged as cloud or cloud shadow and calculated the Normalized Difference Water Index (NDWI) [50] for each Landsat scene. To identify seasonally flooded areas, we extracted areas where the median NDWI during the end of the wet season (September - October) was above zero and the median NDWI during dry season (January – April) was below zero. We then calculated the distance of each pixel to the nearest seasonally flooded pixel, expressed as cumulative cost distance of the shortest path measured in meters from the neatest water source. This was done for each year separately to account for inter-annual variation in flooding extents. Additionally, we added “year” as a continuous variable, to capture inter-annual trends of malaria proportion that are not related to environmental conditions.

S*tatic variables* measured environmental conditions that were not expected to change substantially between years. These included data on settlement structures, land cover and the hydrological setting of the landscape. We created two settlement density indices from 3 m spatial resolution PlanetScope imagery [51]. We acquired images from November 2016 and classified each tile into building and non-building areas with a Random Forest model using a bag fraction of 0.67 over 500 classification trees. The over-all accuracy of the building classification based on out-of-bag data was 0.98 for both areas. In Mecha and Bahir Dar Zuryia we reached a sensitivity of 0.97 and a specificity of 0.98. In Aneded and Awabel sensitivity and specificity were both 0.98. To create variables for settlement density, we resampled the classification to 100 m pixels via a majority filter and then applied a Kernel density estimator using a gaussian kernel with a radius of 100 m and a sigma of 50 m. For each kebele we summarized the mean and maximum settlement density. Settlement classification and density estimation were performed in Google Earth Engine [52].

Land cover variables were derived from a 30 m resolution Landsat-based classification done by Midekisa et al. [24]. We calculated the percentage of each kebele covered by the following land cover classes: open

water, herbaceous wetlands, woody vegetation, cropland, and sparse vegetation.

To map areas that are likely be flooded during larger rain events, we calculated the height above nearest drainage (HAND). The height above nearest drainage measures the vertical distance between a given point and the nearest stream. Low values indicate floodplains and other low-lying areas that are likely to be inundated during and after the rainy season when flow rates are high. We used topographic and stream network data to calculate the HAND using Topography Tools for ArcGIS [53]. We then identified areas less than 2 m above the nearest drainage and calculated the proportion of each kebele that falls into this category.

Statistical analysis

We used Boosted Regression Trees (BRT) to model the relationships between malaria cases and environmental variables. BRT models are commonly used in species distribution modeling [54,55], and have also been used to study mosquito borne diseases [18,56–58]. Advantages of the BRT method are that it performs well under moderate collinearity of predictor variables [59], captures non-linear relationships between predictor and response variables, and allows for interactions between predictor variables. BRT models were created using the R library gbm [60].

For each study area, separate models were fitted for *P. falciparum* + mixed cases, *P. vivax* cases, and total cases. Outpatient numbers were included as offset term. The models were fitted with a learning rate of 0.01, a tree complexity of 3, and a bag fraction of 0.5. Several parameter combinations were tested and the combination that yielded the best R^2 from a 5-fold cross validation was selected.

Relative importance of predictor variables was estimated to determine which variables had the strongest influences on malaria patterns. Variable importance is a measure of how often a variable is used to create a split, normalized by the improvement in squared error resulting from the corresponding split. We then ranked the variables by their importance value and identified the top five variables for each model. To visualize the relationships between these variables and malaria proportion, we fitted partial dependence plots that show how a response variable depends on the predictor variable after taking the average effects of all other variables into account.

Results

Spatio-temporal patterns in malaria cases

Between January 2014 and December 2017, a total of 22,584 malaria cases were reported (Table 2). 59% of all cases were attributed to *P. falciparum* + mixed, and 41% of cases were due to *P. vivax*. In Bahir Dar Zuryia and Mecha, *P. falciparum* + mixed cases made up the largest share of infections, whereas *P. vivax* was more dominant in Aneded and Awabel. Of the four woredas, the highest malaria proportion was in Awabel with 62 malaria cases per 1000 outpatients. A total of 2108 cases had a travel history, meaning

they left the village for at least one night within the last 30 days. The highest proportion of traveled cases was in Awabel with 6 traveled cases per 1000 outpatients.

Table 2
Total malaria case numbers, as well as case number by species, from 2014–2017.

	total malaria cases	*P. falciparum* cases	*P. vivax* cases	mixed cases	traveled cases	outpatients
Aneded	3993	1610	1924	459	201	79394
Awabel	6110	2244	2674	1192	592	97052
Bahir Dar Zuryia	6145	3301	2378	466	257	324769
Mecha	6336	3108	2361	867	1058	416009
The number of traveled cases was also reported as a subset of total malaria cases.						

From 2014 to 2017, we identified a clear downward trend in total malaria cases for Mecha and Bahir Dar Zuryia (Fig. 3). Case numbers in Aneded and Awabel remained more stable over the study period. This led to case numbers in Mecha and Bahir Dar Zuriya being higher than those in Aneded and Awabel during 2014 and 2015, but lower in 2016 and 2017. Both study areas showed strong seasonality with low case occurrence between January – April, a small case peak during May – August and most cases occurring during September – December.

There was considerable variation in total malaria proportion within wordas, as well as for cases broken up into the two pathogen groups *P. falciparum* + mixed and *P. vivax* (Fig. 4). In Mecha and Bahir Dar Zuryia, there was a high malaria proportion in the hilly southern part of the study area, as well as in the flatter central area. In Aneded and Awabel there was a much stronger spatial gradient in malaria proportion with southern kebeles on the Blue Nile escarpment showing a higher malaria proportion. In Mecha and Bahir Dar Zuryia, *P. vivax* was more confined to the kebeles in the southern hills, whereas *P. falciparum* cases were largely responsible for the central cluster. In Aneded and Awabel *P. falciparum* malaria proportion was highest in the kebeles along the Blue Nile escarpment in the south, whereas infections from *P. vivax* cases had additional clusters in the northern and western kebeles. Most cases with a travel history were recorded in the southern hills of Mecha and on the Blue Nile escarpment in Aneded and Awabel.

These general patterns were consistent with the results from the SaTScan analysis (Fig. 5). The kebeles in southern and central Mecha and Bahir Dar Zuryia that showed a high malaria proportion over the entire time period were identified as stable malaria hotspots with malaria clusters in at least 3 years. Similarly, in Aneded and Awabel, areas with overall high malaria proportion were found to be stable malaria hotspots with clusters identified for every year. The SaTScan results show that the high malaria proportion in the southern kebeles of both zones are due to consistent annually recurring outbreaks and not due to individual large outbreaks. Notably, in Aneded and Awabel most kebeles were either identified

as stable hotspots for *P. falciparum* with clusters in three or four years, or as areas that were never identified as hotspots. Only very few kebeles were identified as unstable *P. falciparum* hotspots with clusters in one or two years. *P. vivax* case clusters in Aneded and Awabel were less stable with a considerable number of kebeles being identified as clusters in one or two years.

Association between malaria and environmental variables

We compared model performance metrics for BRT models for each study area and malaria species, as well as total cases. The cross-validated R^2 values were higher in Aneded and Awabel with 0.90 for total cases, 0.91 for *P. falciparum* and mixed cases, and 0.78 for *P. vivax* cases. In Mecha and Bahir Dar Zuryia, the cross-validated R^2 values were 0.6 for total cases, 0.54 for *P. falciparum* and mixed cases and 0.68 for *P. vivax* cases. In Mecha and Bahir Dar Zuryia, the cross-validated R^2 for *P. vivax* was higher than for *P. falciparum*, whereas in Aneded and Awabel the the cross-validated R^2 for *P. falciparum* was higher than for *P. vivax*.

We quantified the relative importance of all variables in the different BRT models and visualized the five most important variables and the respective partial dependence plots to interpret the influence of environmental variables on malaria proportion (Fig. 6). The most important variables differed between the two study areas with only NDVI during the transition season appearing in the top five of both study areas and both malaria species. In both study areas dynamic variables related to vegetation or moisture indices were important, as well as was at least one settlement density index. Except for % Woody vegetation, none of the land cover variables contributed substantially to any of the models. None of the hydrological variables, like the distance to seasonal water, height above nearest drainage, or percent of land within the Koga irrigation scheme, were among the most important variables.

In Mecha and Bahir Dar Zuryia, *P. falciparum* was most strongly influenced by NDVI during the transition season (10 %), with higher malaria proportion coinciding with lower NDVI values. Similarly, lower NDMI values during transition season were also associated with a higher malaria proportion. Precipitation during rainy season was associated with higher malaria proportion. However, all variables were at or below 10 % in variable importance. This result, and the fact that the model for *P. falciparum* in Mecha and Bahir Dar Zuryia had the lowest cross-validated R^2, suggest that additional unmeasured factors also influence malaria patterns in this landscape. *P. vivax* in Mecha and Bahir Dar Zuryia was mostly influenced by NDVI during transition season (17 %) with a higher malaria proportion related to lower NDVI values, and maximum settlement index (13 %) with higher cases in kebeles with the largest settlement clusters. Year (8 %) was also an important variable in Mecha and Bahir Dar Zuryia, indicating that the downward trend over the entire study period was driven by processes not related to environmental factors included in this model.

In Aneded and Awabel, NDMI during the dry season was by far the most important variable for both malaria species (41 % for *P. falciparum* and 36 % for *P. vivax*). Additionally, lower NDVI during transmission season was also associated with more malaria for both species (both < 10%). Both spectral

indices showed an association of drier conditions with malaria occurrence. Higher malaria proportion was also associated with percent cover with woody vegetation for *P. falciparum* (11%), and *P. vivax* (6%). Higher LST during transition season had a positive effect on malaria proportion for both species. However, the effect on *P. falciparum* and mixed cases (14%) was larger than the effect on *P. vivax* (6%). For both species, a mean settlement index close to zero was associated with a higher malaria proportion (both < 10%).

Discussion

This study presents an analysis of kebele-level Malaria proportion in the Amhara region in Ethiopia between 2014 and 2017. Remotely sensed environmental variables were used to explore the influences of environmental risk factors,. We found considerable small-scale variation in malaria proportion and malaria case clusters between kebeles of the same study area. Climatic variables, settlement structure, and spectral indices influenced malaria risk, whereas variables related to seasonal waterbodies, flood plains, as well as irrigation showed little effect. However, the importance of the variables, as well as the association between those variables and malaria, varied between the two study areas, as well as between the different *Plasmodium* species. These results highlight the importance of stratifying disease risk models into zones with similar geographic context.

Climatic variables like temperature and precipitation influenced malaria proportion in both landscapes. This result does not come as a surprise, as the influences of temperature and precipitation on malaria are well established for the area [19, 22, 23]. Our study shows that in a geographically diverse area like the Amhara region, climatic variables can vary greatly within woredas and drive sub-woreda level malaria proportion. There were several important differences between the two study areas. Mecha and Bahir Dar Zuryia had a very strong precipitation gradient, and we found precipitation to be an important predictor variable for malaria. Temperature was found to be a stronger predictor of *P. falciparum* in Aneded & Awabel, despite both study areas showing a considerable temperature gradient. Previous studies have found that due to the different biology of the two parasites, *P. falciparum* may be limited to lower elevations with warmer temperatures, whereas *P. vivax* is more tolerant of lower temperatures and therefore more stable across higher elevations in the Amhara region [35, 61]. The effects of temperature on *P. falciparum* in Aneded & Awabel may therefore be due to a temperature threshold that is reached only in the higher elevations of this study area. Overall, these results support the findings of previous studies that there is variation between woredas in how sensitive malaria cases are to climatic conditions [20].

Spectral indices (NDVI and NDMI) explained substantial variation in malaria proportion. However, we found that drier and less green conditions were to be associated with more malaria. Given the importance of water and moisture for the mosquito life cycle, these results appear counter-intuitive. However, the associations between spectral incides and malaria proportion may in part reflect underlying correlations with other aspects of the physical and social environment. In Aneded and Awabel, the driest condition

with the warmest kebeles located on the Blue Nile escarpment. In Mecha and Bahir Dar Zuryia, the highest moisture values were observed in kebeles that are part of the Koga irrigation scheme.

The negative influences of NDVI and NDMI on malaria may reflect socio-economic factors that influence malaria infection more directly. For example, there is a relationship between the risk of getting infected with malaria and the main source of water of a household, with households that need less time to access water being at lower risk than households that require more time to access water [32]. This effects may be enhanced in drier environments with less water availability. The relationship between dry conditions and work-related migration could also play a role. Travel has been shown to be a significant risk factor for malaria transmission particularly for *P. falciparum* related infections [62]. In the Amhara region, high precipitation variability and low net primary production has been shown to increase out-migration [63]. In rural areas in Amhara, where seasonal migrants often leave their village to seek an additional income elsewhere [64], drier conditions may lead to increased seasonal migration and risk of malaria importation. From our own data, we have evidence that those dry areas with high malaria proportion coincide with those areas where a large proportion of the patients had a travel history. However, from the available data we could not derive which *Plasmodium* species the patients with travel history were infected with, nor could we tell where and for how long they traveled. We could therefore not include this information into our formal analysis, but we hypothesize that the socio-economic factors that are associated with dry conditions could increase travel and imported malaria in some situations.

Settlement structure was found to be an important variable in all models, and the relationship between settlement index and malaria proportion differed between the two study areas. In Mecha and Bahir Dar Zuryia malaria proportion was higher in kebeles with a high maximum settlement index. This included kebeles with a dense settlement cluster, regardless of the amount of agricultural land in an area. In Aneded and Awabel we found the highest malaria proportion in kebeles with a mean settlement density index close to zero, which was found in large and very sparsely populated kebeles. The relationship between low mean settlement density and high malaria proportion in Aneded and Awabel may be related to previously discussed environmental factors. The sparsely populated kebeles on the escarpment are the warmest and driest areas in Aneded and Awabel. Those kebeles are also very large and in a very steep landscape. Individual households within those kebeles are geographically more isolated from the next village or town than households anywhere else in the study area, and may therefore have difficulties in accessing to resources known to reduce infection risk, such as information about malaria and fever treatment [65], or malaria prevention tools [66]. In Mecha and Bahir Dar Zuryia, households in even the most rural kebeles were geographically less isolated, which is one reason why we might see a different relationship between malaria proportion and settlement structure.

Water-related variables did not influence malaria proportion in our model. These included the coverage of open water, herbaceous wetlands, percent of land within the Koga irrigation scheme, the average distance to the nearest seasonal waterbody, as well as areas likely to flood due to their height above the nearest drainage. These findings are in contrast to the general expectation of relationships between malaria cases and temporary water bodies. Other studies from different parts of Ethiopia have found that

woredas with high wetlands cover had high malaria incidence [24] and distance to breeding sites for samples of individual households influences malaria risk on household level [23, 67, 68]. However, those previous studies have been performed at different spatial scales, and the proximity of temporary water bodies was much less important at the kebele level within he woredas that we studied.

In particular, we expected to find higher malaria proportion in kebeles located within the Koga irrigation scheme, as irrigation has been shown to created breeding habitat for anopheline mosquitoes, and effects were previously observed for large dam projects in Ethiopia [69]. The lack of an irrigation effect on malaria in our study may be an example of what Ijumba and Lindsay describe as the “paddies paradox” [70]. Because the Koga irrigation schemes increased wealth in the irrigator households [37], malaria case numbers may be low despite the fact that new irrigation channels can act as potential breeding sites for mosquitoes. The new socio-economic structure with increased wealth, and better access to health care and education may contribute to a low burden of malaria in this region.

The utilization of remotely sensed satellite data to study malaria infection patterns comes with limitations. The nature of satellite Earth observation data allows us to study only the physical properties of the earth surface. We therefore could not directly measure socio-economic factors like seasonal migration patterns education or access to health care. Particularly in Mecha and Bahir Dar Zuryia, the cross validated R^2 of 0.54 and 0.68 for models of malaria cases indicate that environmental variables alone cannot explain the variation in malaria proportion. This is also reflected in the strong downward trend of malaria cases due to the effectiveness of recent intervention programs. However, as this study primarily focused on leveraging different sources of remotely-sensed environmental factors, the influence of socio-economic factors, as well as the influence of intervention programs was beyond our scope. We faced additional limitations in classifying settlements via PlanetScope imagery. The spatial resolution of 3 m was sufficient to capture most buildings but was not sufficient to capture small huts with thatched roofs. To capture these buildings, imagery with an even higher spatial resolution would be necessary.

Conclusion

Relationships between spatio-temporal patterns of malaria proportion and environmental variables derived from satellite imagery varied in two different landscapes in Ethiopia, and were different from results of previous malaria-environment studies conducted at coarser woreda resolution and finer household resolutions. Relationships between climatic variables and malaria proportion followed the expected pattern with higher temperatures and more rainfall leading to a higher malaria proportion. Other associations did not fit our general expectation of increased malaria in rural areas with more temporary waterbodies. Instead we found contrasting relationships between settlement density and malaria in the two study areas, as well as an association between malaria and dry landscapes in both study areas. The relationships of these environmental variables with malaria likely reflected indirect relationships with socio-economic factors such as seasonal migration, water management, and access to health care that affect malaria risk. Recent interventions have lowered malaria incidence and may also have modified some of these malaria-environment associations. These results emphasize that studies of malaria-

environmental relationships and predictive models of malaria occurence should be context specific to account for such differences.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Satellite data products MODIS, Landsat, and IMGERG data used are publicly available and were accessed via Google Earth Engine (https://code.earthengine.google.com/). PlanetScope data are accessible for researchers via the “Education and Research Program (https://www.planet.com/markets/education-and-research/).

Public health data that support the findings of this study are not publicly available, as they were obtained via a data sharing agreement with the Amhara Regional Health Bureau.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been supported by the National institute of Allergy and Infectious Diseases (R01-AI079411)

Authors contributions

AnM and MCW conceptualized the study, AAA, ML, WA, AbM acquired the data and contributed to the interpretation of the data. AnM performed the the main analysis and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Justin K. Davis and Dr. Dawn M. Nekorchuk for providing technical help with data analysis.

References

1. Tanner M, Greenwood B, Whitty CJM, Ansah EK, Price RN, Dondorp AM, et al. Malaria eradication and elimination: views on how to translate a vision into reality. BMC Med. 2015;13:1–22.
2. Newby G, Bennett A, Larson E, Cotter C, Shretta R, Phillips AA, et al. The path to eradication: a progress report on the malaria-eliminating countries. Lancet. 2016;387:1775–84.
3. WHO. World malaria report 2019. Geneva; 2019.
4. United Nations. The Millennium Development Goals Report [Internet]. New York; 2013. Available from: https://www.un.org/millenniumgoals/pdf/report-2013/mdg-report-2013-english.pdf
5. Pigott DM, Atun R, Moyes CL, Hay SI, Gething PW. Funding for malaria control 2006–2010: A comprehensive global assessment. Malar J. 2012;11.
6. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS Med. 2012;9:e1001165–e1001165.
7. Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, Bhatt S, et al. Global mapping of infectious disease. Philos Trans R Soc B Biol Sci. 2013;368:20120250–20120250.
8. Pigott DM, Howes RE, Wiebe A, Battle KE, Golding N, Gething PW, et al. Prioritising infectious disease mapping. PLoS Neglected Trop Dis. 2015;9:e0003756–e0003756.
9. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. Ecol Lett. 2019;22:1690–708.
10. Bayoh MN. Studies on the development and survival of Anopheles gambiae sensu stricto at various temperatures and relative humidities [Doctoral]. Durham University; 2001.
11. Munga S, Minakawa N, Zhou G, Mushinzimana E, Barrack O-OJ, Githeko AK, et al. Association between land cover and habitat productivity of malaria vectors in western Kenyan highlands. Am J Trop Med Hyg. 2006;74:69–75.
12. Hardy AJ, Gamarra JGP, Cross DE, Macklin MG, Smith MW, Kihonda J, et al. Habitat hydrology and geomorphology control the distribution of malaria vector larvae in rural Africa. PLOS ONE. 2013;8:e81931.
13. Kibret S, Wilson GG, Tekie H, Petros B. Increased malaria transmission around irrigation schemes in Ethiopia and the potential of canal water management for malaria vector control. Malar J. 2014;13:360.
14. Kabaria CW, Molteni F, Mandike R, Chacky F, Noor AM, Snow RW, et al. Mapping intra-urban malaria risk using high resolution satellite imagery: a case study of Dar es Salaam. Int J Health Geogr. 2016;15:26.

15. Kazansky Y, Wood D, Sutherlun J. The current and potential role of satellite remote sensing in the campaign against malaria. Acta Astronaut. 2016;121:292–305.
16. Parselia E, Kontoes C, Tsouni A, Hadjichristodoulou C, Kioutsioukis I, Magiorkinis G, et al. Satellite earth observation data in epidemiological modeling of malaria, dengue and West Nile virus: a scoping review. Remote Sens. Multidisciplinary Digital Publishing Institute; 2019;11:1862.
17. Ebhuoma O, Gebreslasie M. Remote sensing-driven climatic/environmental variables for modelling malaria transmission in sub-saharan Africa. Int J Environ Res Public Health. Multidisciplinary Digital Publishing Institute; 2016;13:584.
18. Solano-Villarreal E, Valdivia W, Pearcy M, Linard C, Pasapera-Gonzales J, Moreno-Gutierrez D, et al. Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon. Sci Rep. Nature Publishing Group; 2019;9:15173.
19. Midekisa A, Senay G, Henebry GM, Semuniguse P, Wimberly MC. Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. Malar J. 2012;11:165–165.
20. Davis JK, Gebrehiwot T, Worku M, Awoke W, Mihretie A, Nekorchuk D, et al. A genetic algorithm for identifying spatially-varying environmental drivers in a malaria time series model. Environ Model Softw. 2019;119:275–84.
21. Machault V, Vignolles C, Borchi F, Vounatsou P, Pages F, Briolant S, et al. The use of remotely sensed environmental data in the study of malaria. Geospatial Health. 2011;5:151–68.
22. Minale AS, Alemu K. Malaria-risk assessment using geographical information system and remote sensing in Mecha District, West Gojjam, Ethiopia. Geospatial Health. 2018;13:157–63.
23. Belay DB, Kifle YG, Goshu AT, Gran JM, Yewhalaw D, Duchateau L, et al. Joint Bayesian modeling of time to malaria and mosquito abundance in Ethiopia. BMC Infect Dis. 2017;17:415–415.
24. Midekisa A, Senay GB, Wimberly MC. Multisensor earth observations to characterize wetlands and malaria epidemiology in Ethiopia. Water Resour Res. 2014;50:8791–806.
25. Stresman GH. Beyond temperature and precipitation: Ecological risk factors that modify malaria transmission. Acta Trop. 2010;116:167–72.
26. Olson SH, Gangnon R, Elguero E, Durieux L, Guégan JF, Foley JA, et al. Links between climate, malaria, and wetlands in the amazon basin. Emerg Infect Dis. 2009;15:659–62.
27. Midekisa A, Beyene B, Mihretie A, Bayabil E, Wimberly MC. Seasonal associations of climatic drivers and malaria in the highlands of Ethiopia. Parasit Vectors. 2015;8:339.
28. Hamm NAS, Soares Magalhães RJ, Clements ACA. Earth observation, spatial data quality, and neglected tropical diseases. Lv S, editor. PLoS Negl Trop Dis. 2015;9:e0004164.
29. Lankir D, Solomon S, Gize A. A five-year trend analysis of malaria surveillance data in selected zones of Amhara region, Northwest Ethiopia. BMC Public Health. 2020;20:1175.
30. Lemma W. Impact of high malaria incidence in seasonal migrant and permanent adult male laborers in mechanized agricultural farms in Metema – Humera lowlands on malaria elimination program in Ethiopia. BMC Public Health. 2020;20:320.

31. Merkord CL, Liu Y, Mihretie A, Gebrehiwot T, Awoke W, Bayabil E, et al. Integrating malaria surveillance with climate data for outbreak detection and forecasting: the EPIDEMIA system. Malar J. 2017;16:89.
32. Ayele DG, Zewotir TT, Mwambi HG. Prevalence and risk factors of malaria in Ethiopia. Malar J. 2012;11:195–195.
33. Esayas E, Woyessa A, Massebo F. Malaria infection clustered into small residential areas in lowlands of southern Ethiopia. Parasite Epidemiol Control. 2020;10:e00149.
34. Zemene E, Koepfli C, Tiruneh A, Yeshiwondim AK, Seyoum D, Lee M-C, et al. Detection of foci of residual malaria transmission through reactive case detection in Ethiopia. Malar J. 2018;17:390.
35. Yalew WG, Pal S, Bansil P, Dabbs R, Tetteh K, Guinovart C, et al. Current and cumulative malaria infections in a setting embarking on elimination: Amhara, Ethiopia. Malar J. 2017;16:242.
36. Scott CA, Yeshiwondim AK, Serda B, Guinovart C, Tesfay BH, Agmas A, et al. Mass testing and treatment for malaria in low transmission areas in Amhara Region, Ethiopia. Malar J. 2016;15:305.
37. Kassie KE, Alemu BA, Wedajoo AS. Impact of irrigation on household multidimensional poverty reduction in the Koga irrigation development project, northern Ethiopia. Asian Dev Perspect. 2018;29.
38. Abeku TA, Vlas SJD, Borsboom G, Teklehaimanot A, Kebede A, Olana D, et al. Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best. Trop Med Int Health. 2002;7:851–7.
39. Animut A, Balkew M, Lindtjørn B. Impact of housing condition on indoor-biting and indoor-resting Anopheles arabiensis density in a highland area, central Ethiopia. Malar J. 2013;12:393.
40. Federal Ministry of Health. National malaria elimination strategic plan: 2021-2025. Addis Ababa; 2020.
41. Guintran J-O, Delacollette C, Trigg P. Systems for the early detection of malaria epidemics in Africa: an analysis of current practices and future priorities. Geneva: World Health Organization; 2006.
42. Wimberly MC, Midekisa A, Semuniguse P, Teka H, Henebry GM, Chuang T, et al. Spatial synchrony of malaria outbreaks in a highland region of Ethiopia. Trop Med Int Health. 2012;17:1192–201.
43. Kulldorff M. SaTScan TM User Guide. 2018.
44. Wan Z, Hook S, Hulley G. MOD11A2 MODIS/Terra land surface temperature/emissivity 8-day L3 global 1km SIN grid V006. NASA EOSDIS Land Processes DAAC; 2015.
45. Schaaf C, Wang Z. MCD43C4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF-Adjusted Ref Daily L3 Global 0.05Deg CMG V006 [Data set]. 2015;
46. Tucker CJ. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens Environ. 1979;8:127–50.
47. Gao BC. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ. 1996;58:257–66.
48. Venables W, Ripley B. Modern Applied Statistics with S. 4th ed. New York: Springer; 2002.
49. Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, et al. The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at

fine scales. J Hydrometeorol. American Meteorological Society; 2007;8:38–55.
50. McFeeters SK. The use of the normalized difference water index (NDWI) in the delineation of open water features. Int J Remote Sens. 1996;17:1425–32.
51. Planet Team. Planet application program interface: in space for life on earth. San Francisco, CA; 2017.
52. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens Environ. 2017;202:18–27.
53. Dilts TE. Topography Tools for ArcGIS 10.1. University of Nevada Reno; 2015.
54. Elith J, Graham CH. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography. 2009;32:66–77.
55. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008;77:802–13.
56. Hess A, Davis JK, Wimberly MC. Identifying environmental risk factors and mapping the distribution of West Nile virus in an endemic region of North America. GeoHealth. 2018;2:395–409.
57. Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. Nat Microbiol. 2019;1.
58. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496:504–504.
59. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 2013;36:27–46.
60. Ridgeway G, Southworth M, RUnit. Package 'gbm.' CRAN; 2013.
61. Lyon B, Dinku T, Raman A, Thomson MC. Temperature suitability for malaria climbing the Ethiopian Highlands. Environ Res Lett. 2017;12:064015–064015.
62. Yukich JO, Taylor C, Eisele TP, Reithinger R, Nauhassenay H, Berhane Y, et al. Travel history and malaria infection risk in a low-transmission setting in Ethiopia: a case control study. Malar J. 2013;12:33.
63. Hermans-Neumann K. Human migration, climate variability, and land degradation: hotspots of socio-ecological pressure in Ethiopia. Reg Env Change. 2017;17:1479–92.
64. Asfaw W, Tolossa D, Zeleke G. Causes and impacts of seasonal migration on rural livelihoods: Case studies from Amhara Region in Ethiopia. Nor Geogr Tidsskr. 2010;14.
65. Hwang J, Graves PM, Jima D, Reithinger R, Kachur SP, Group and the EMW. Knowledge of malaria and its association with malaria-related behaviors—results from the malaria indicator survey, Ethiopia, 2007. PLOS ONE. Public Library of Science; 2010;5:e11692.
66. Paulander J, Olsson H, Lemma H, Getachew A, Sebastian MS. Knowledge, attitudes and practice about malaria in rural Tigray, Ethiopia. Glob Health Action. Taylor & Francis; 2009;2:1839.
67. Alemu A, Tsegaye W, Golassa L, Abebe G. Urban malaria and associated risk factors in Jimma town, south-west Ethiopia. Malar J. 2011;10:173.

68. Nissen A, Cook J, Loha E, Lindtjørn B. Proximity to vector breeding site and risk of Plasmodium vivax infection: a prospective cohort study in rural Ethiopia. Malar J. 2017;16:380.
69. Kibret S, Wilson GG, Ryder D, Tekie H, Petros B. Malaria impact of large dams at different eco-epidemiological settings in Ethiopia. Trop Med Health. 2017;45:4–4.
70. Ijumba JN, Lindsay SW. Impact of irrigation on malaria in Africa: paddies paradox. Med Vet Entomol. 2001;15:1–11.

Figures

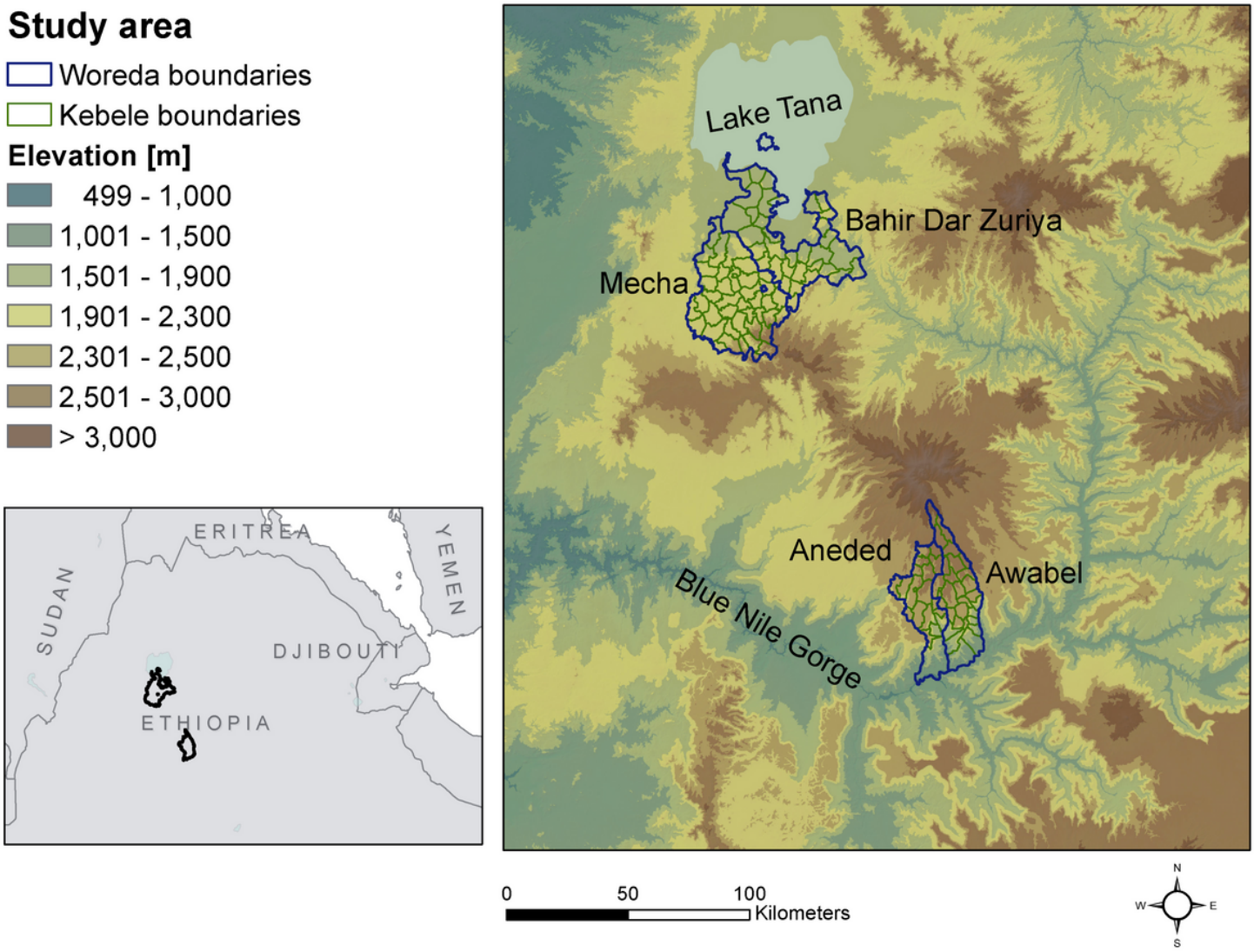

Figure 1

Study area in the Amhara region of Ethiopia. Blue lines indicate woredas (district) boundaties, green lines indicate the kebeles within those woredas.

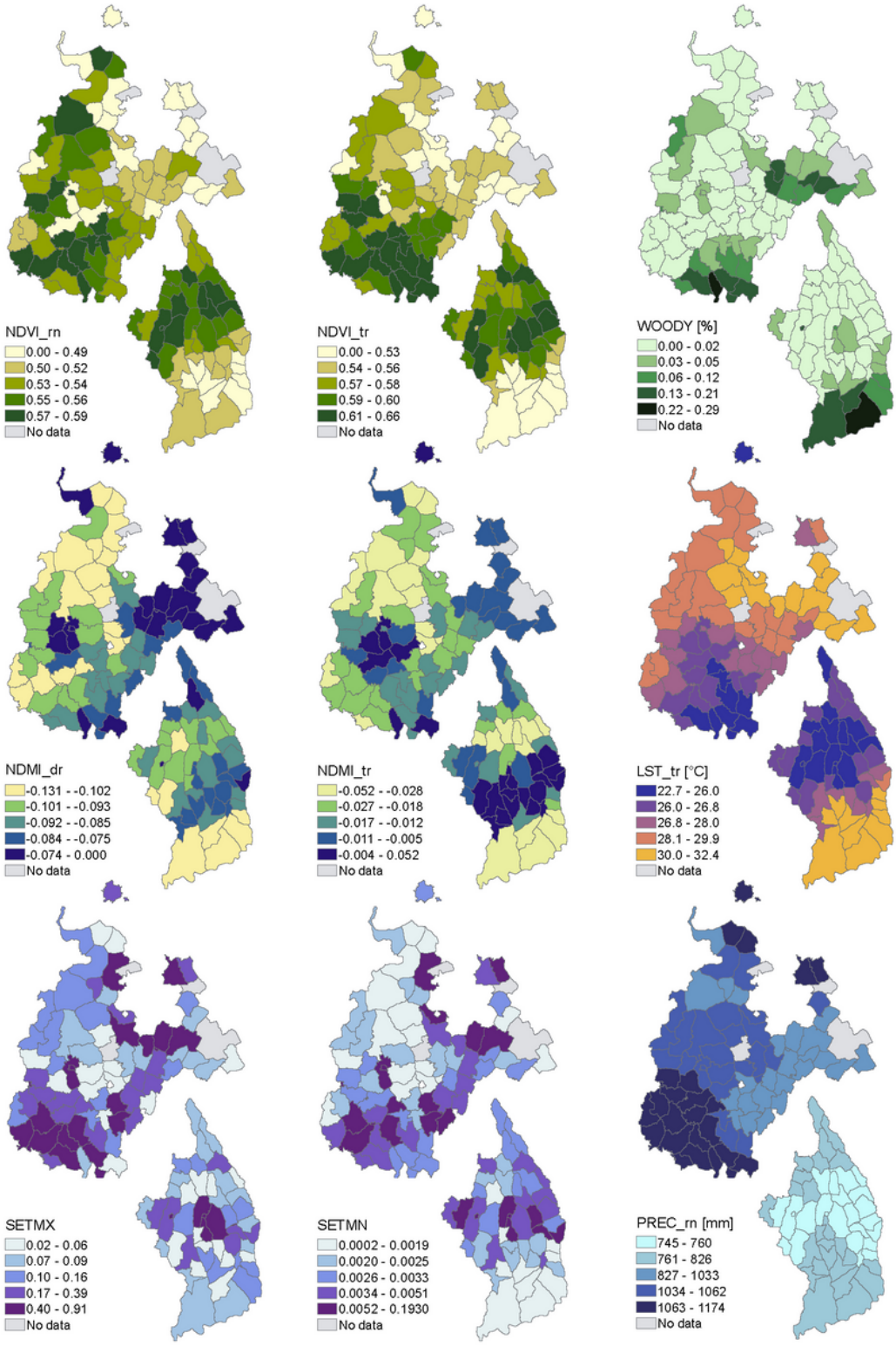

Figure 2

Maps displaying the variation in environmental conditions within the two study areas. See variable name abbreviations in Table 1. Note that the positions of the two study areas do not represent their true proximity. See the map in Figure 1 for their actual locations.

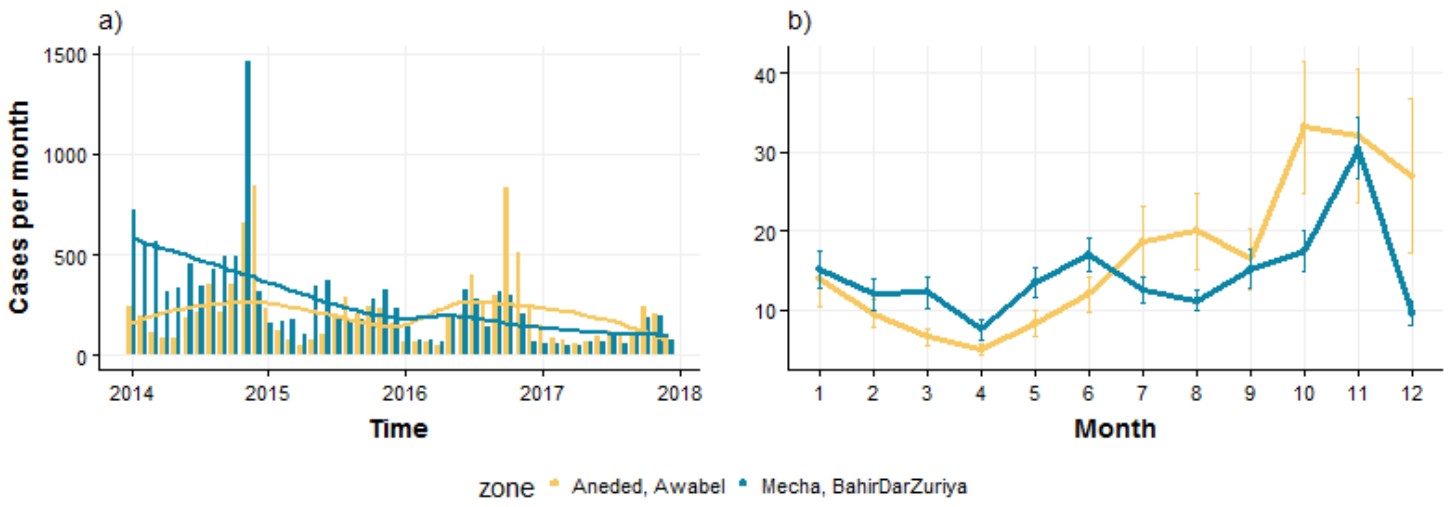

Figure 3

Total malaria case numbers from Jan 2014 – Dec 2017 by study area. a) Monthly case numbers shown over entire study period. Lines indicate LOESS smoothed trend. b) Seasonality of total malaria cases. Lines represent mean case numbers over all kebeles and years. Error bars indicate the standard error of the observed mean.

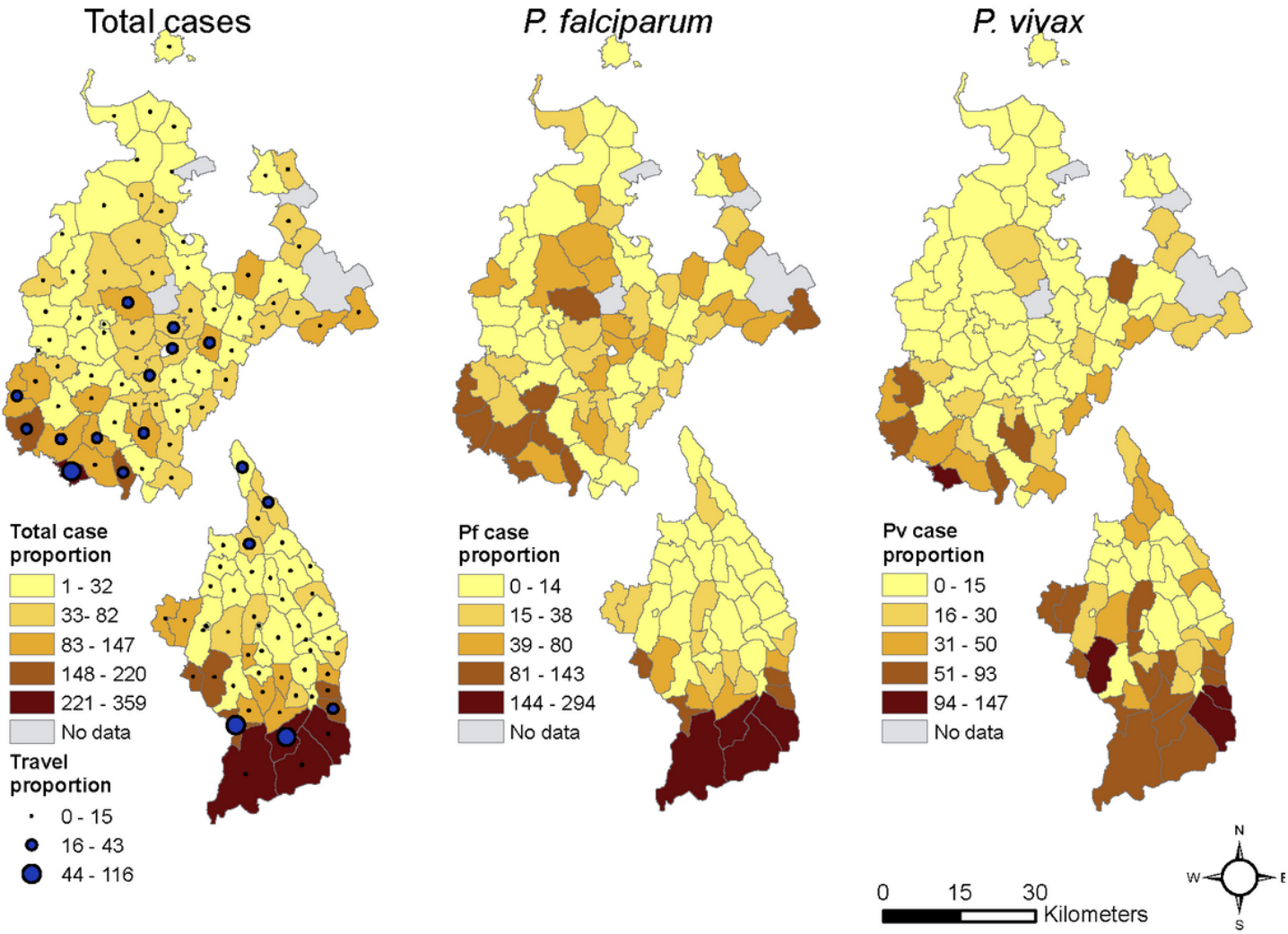

Figure 4

Spatial distribution of malaria proportion mapped are malaria cases per 1000 outpatients for total malaria cases, P. falciparum + mixed cases, and P. vivax cases. The upper maps are the woredas Mecha & Bahir Dar Zuryia, and the lower maps are the woredas Aneded & Awabel. Blue dots indicate the proportion of total malaria patients with travel history per 1000 outpatients.

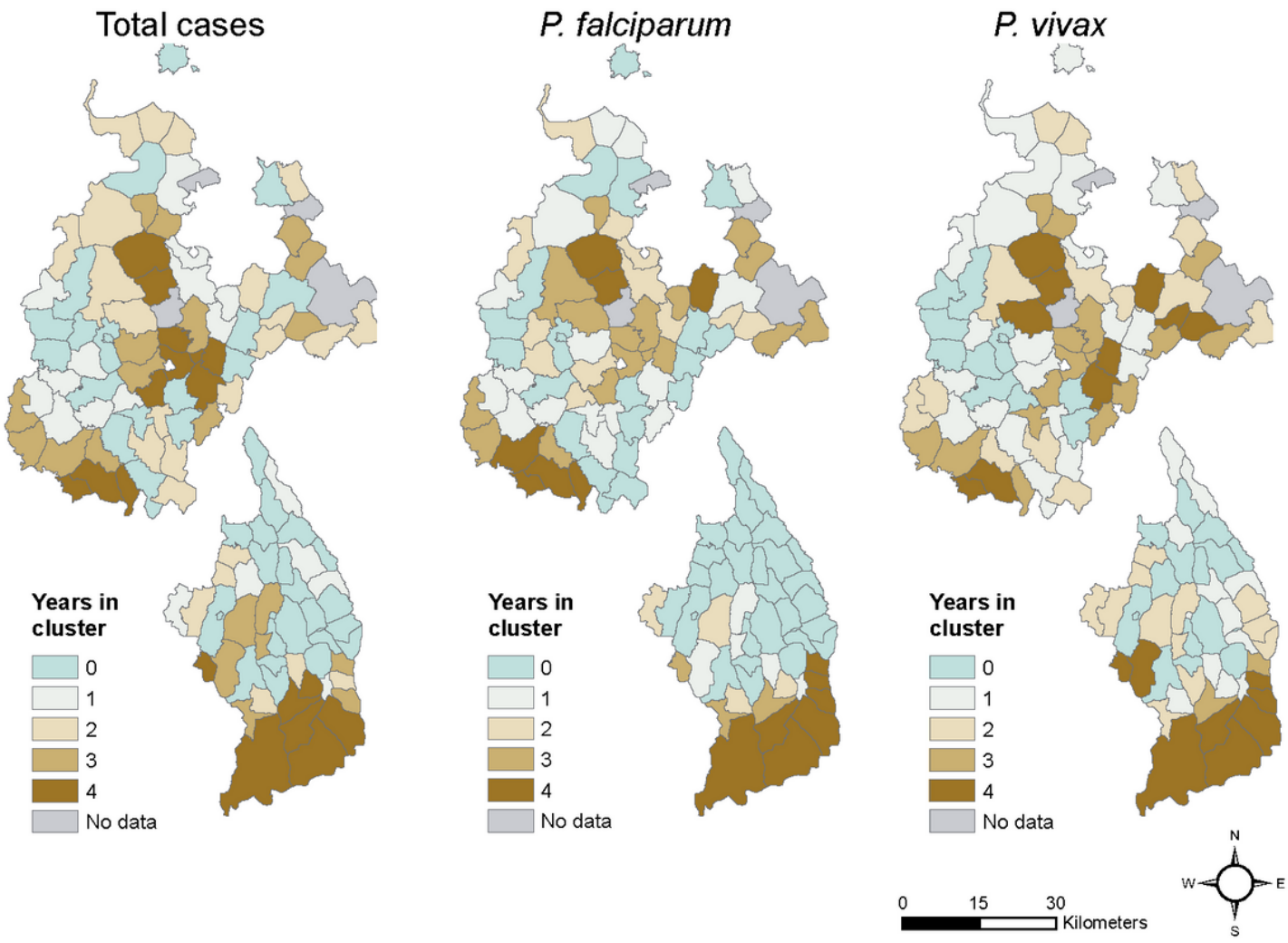

Figure 5

Number of years each kebele was identified as part of a case cluster by the SaTScan analyses.

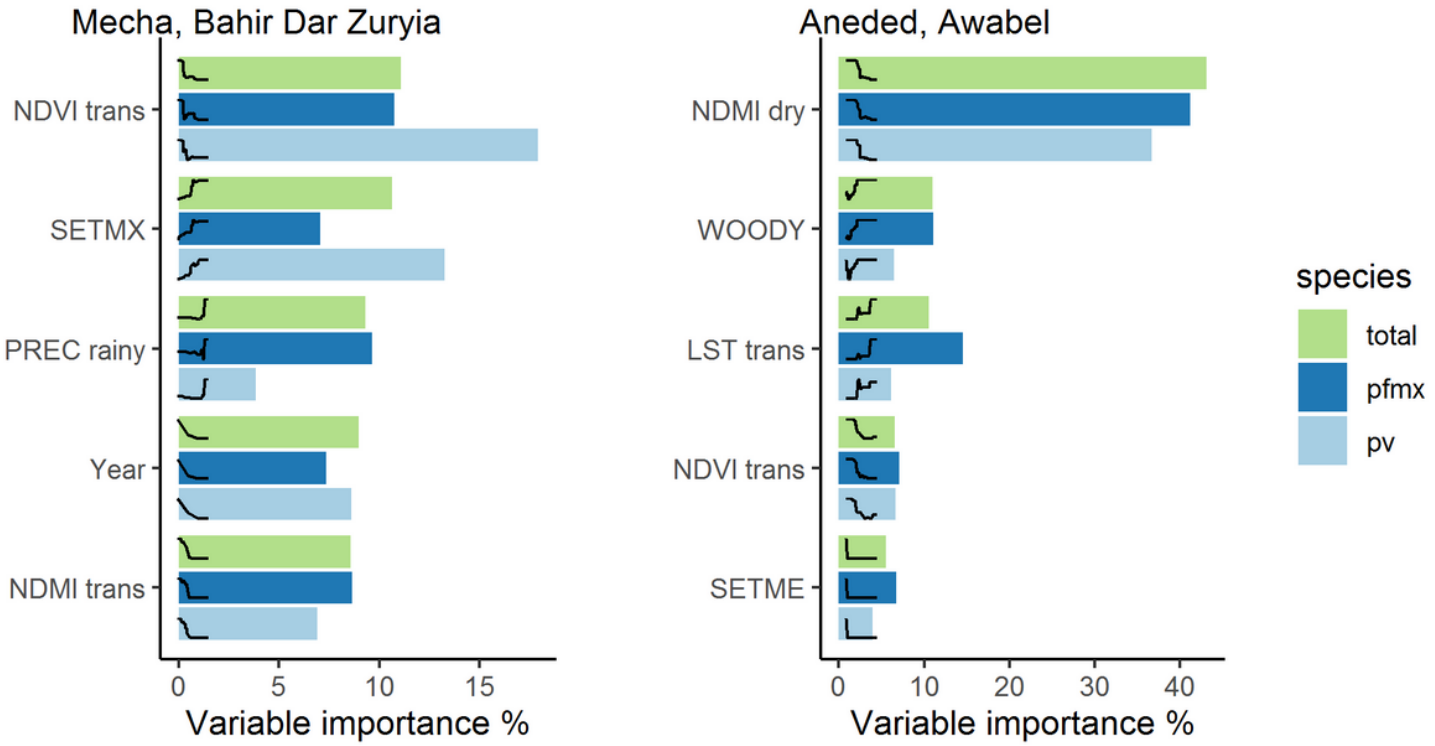

Figure 6

Variable importance of the top 5 environmental variables as identified by boosted regression tree models. Different color bars represent different malaria species summaries (green = total species, dark blue = P. falciparum and mixed infections, light blue = P. vivax). Curves represent the fitted partial dependence curve for each variable.