**Multisensor earth observations to characterize wetlands and malaria epidemiology in Ethiopia**

Alemayehu Midekisa¹, Gabriel B. Senay², and Michael C. Wimberly¹

¹Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, South Dakota, USA, ²U.S. Geological Survey Earth Resources Observation and Science Center, Sioux Falls, South Dakota, USA

**Abstract** Malaria is a major global public health problem, particularly in Sub-Saharan Africa. The spatial heterogeneity of malaria can be affected by factors such as hydrological processes, physiography, and land cover patterns. Tropical wetlands, for example, are important hydrological features that can serve as mosquito breeding habitats. Mapping and monitoring of wetlands using satellite remote sensing can thus help to target interventions aimed at reducing malaria transmission. The objective of this study was to map wetlands and other major land cover types in the Amhara region of Ethiopia and to analyze district-level associations of malaria and wetlands across the region. We evaluated three random forests classification models using remotely sensed topographic and spectral data based on Shuttle Radar Topographic Mission (SRTM) and Landsat TM/ETM+ imagery, respectively. The model that integrated data from both sensors yielded more accurate land cover classification than single-sensor models. The resulting map of wetlands and other major land cover classes had an overall accuracy of 93.5%. Topographic indices and subpixel level fractional cover indices contributed most strongly to the land cover classification. Further, we found strong spatial associations of percent area of wetlands with malaria cases at the district level across the dry, wet, and fall seasons. Overall, our study provided the most extensive map of wetlands for the Amhara region and documented spatiotemporal associations of wetlands and malaria risk at a broad regional level. These findings can assist public health personnel in developing strategies to effectively control and eliminate malaria in the region.

**1. Introduction**

Malaria is a major global public health problem with an estimated 207 million cases and 627,000 deaths annually, where the majority of these deaths are in sub-Saharan Africa [WHO, 2013]. The disease is caused by *Plasmodium* parasites transmitted to and acquired from humans by female mosquitoes during blood meals. Spatial heterogeneity in malaria transmission has been reported in previous research due to variability in hydrological processes, physiography, land cover, and social factors [Bousma et al., 2012; Gaudart et al., 2006; Kelly et al., 2012; Kreulets et al., 2008]. One important landscape feature that influences heterogeneity in mosquito habitats is the spatial pattern of wetlands, which can be a potential foci for malaria transmission. For example, herbaceous wetlands can sustain malaria transmission during the dry season and then facilitate expansion into the broader landscape during the wet season [Bousma et al., 2012]. As a result, mapping and monitoring of hot spots such as herbaceous wetlands may help to target interventions such as long-lasting insecticide-treated nets, indoor residual spraying, larval source management, and other public health interventions to control and eliminate malaria.

The influences of environmental risk factors on malaria parasite development are well established. Temperature influences parasite development inside the female mosquito as well as mosquito development and survival [Lyimo et al., 1992; Martens et al., 1995; Paaijmans et al., 2010]. In addition, the aquatic developmental cycle of mosquitoes, including egg, larva, and pupa stages, requires the availability of standing water [Martens et al., 1995]. Precipitation thus affects malaria transmission because it is the ultimate source of water that is needed to create these aquatic habitats [Hardy et al., 2013; Martens et al., 1999]. However, the relationship between malaria and precipitation is poorly understood, and previous research has shown variable associations of precipitation with malaria transmission in different geographic settings [Hardy et al., 2013; Midekisa et al., 2012; Paaijmans et al., 2007; Zhang et al., 2008]. The effect of rainfall on vector habitat formation is complicated by other factors such as physiographic and land cover characteristics that influence
hydrological processes and thereby constrain the availability of surface water for malaria vector breeding [Smith et al., 2013].

Previous studies have shown evidence that local variation in hydrological processes, physiography, and land cover will influence the aquatic habitats for vector mosquitoes. For example, a study in the Kenyan highlands reported that topographic wetness was highly associated with the spatial distribution of malaria cases [Cohen et al., 2010]. Another study in northwest Pennsylvania, USA emphasized that semipermanent wetlands were major potential larval habitats for the vector mosquitoes [Chase and Knight, 2003]. Larvae of Anopheles arabiensis were identified in riverine sand pools in the middle course of the Ethiopian Rift Valley [Kenea et al., 2011]. In western Kenya, Anopheles gambiae were more abundant in areas with wetter soils, especially in the dry season [Minakawa et al., 2001]. The ecological role of wetlands may be particularly critical in geographic regions where there is seasonal variability in precipitation because they can sustain soil moisture during the dry season [Teferi et al., 2010]. These and other studies have provided evidence that local physiography, hydrological processes, and land cover patterns are key factors that influence the availability of surface water that provides suitable breeding habitats for mosquitoes. However, previous research that assessed the association of wetlands or topography with mosquitoes or malaria has been focused on relatively small landscape (e.g., hundreds to thousands of ha). There is a need for broader regional level studies because this is the level at which public health decisions are implemented.

Earth observation satellites provide frequent measurements of global land surface characteristics. Availability of multisensor and multitemporal measurements from satellites enhances the ability to identify and monitor environmental risk factors of mosquito-borne diseases [Chuang and Wimberly, 2012; Hay et al., 1998, 2000; Kalluri et al., 2007; Machault et al., 2011; Midekisa et al., 2012]. Previous studies have used remote sensing observations to identify mosquito breeding habitats such as natural wetlands [Beck et al., 1994; Brown et al., 2008; Mushinizimana et al., 2006]. High spatial resolution remote sensing data were used to map endemic areas of malaria in Burkina Faso [Dambach et al., 2009]. There has also been recent progress in using multisensor remote sensing data for early warning of vector-borne diseases [Chabot-Couture et al., 2014; Machault et al., 2010; Midekisa et al., 2012]. All these previous studies emphasized the potential of remote sensing to map and monitor environmental risk factors for vector-borne diseases.

One of the strengths of satellite remote sensing is the potential for efficiently mapping environmental risk factors, including wetlands, across large spatial extents. For example, an earlier work in South Sudan used satellite-derived evapotranspiration (ET) to assess the seasonal variation of the Sudd wetlands, which cover more than 30,000 km² [Rebelo et al., 2012]. Wetland maps for the entire Congo Basin were developed using multisource remote sensing data [Bwangoy et al., 2010]. Remote sensing-based wetland maps were also developed for the province of Queensland, Australia [Knight et al., 2009]. More recently, a study in China developed wetland maps for the Yellow River Delta using Landsat data [Liu et al., 2014]. However, these regional wetland maps have not yet been widely utilized in public health applications. Our work builds on earlier studies of malaria-wetland associations and previous broad-scale wetland mapping efforts to assess the spatial associations of malaria and wetlands at a large regional extent. Ultimately, these wetland maps can provide additional sources of information to public health decision makers, which can help them to prioritize the allocation of limited resources to areas with high risk of malaria transmission [Kelly et al., 2012].

The aim of this paper is to quantify the spatial distribution of herbaceous wetlands and test their association with malaria transmission at a regional scale in the Amhara region of Ethiopia. In this malaria epidemic-prone region, herbaceous wetlands are distributed heterogeneously across the landscape. Wetland vegetation is typically dominated by grasses, and wetlands are often used as communal grazing lands. Our overarching hypothesis is that the pattern of wetlands, along with elevation and climatic variables, is a determinant of the regional distribution of malaria incidence. The rationale is that wetlands serve as natural reservoirs for malaria vector mosquitoes, particularly in the dry season, and can also facilitate expansion of malaria transmission throughout the broader region in the wet season [Bousema et al., 2012]. Thus, the objectives of the present study are (1) to develop a land cover classification for the Amhara region with a particular focus on wetland mapping, and (2) to test the district-level spatial associations between herbaceous wetlands and malaria incidence at the regional extent. We achieved these aims by quantifying the spatial distribution of herbaceous wetlands using data from multisensor satellite observations and analyzing their spatial associations with epidemiological surveillance data.
2. Methods

2.1. Study Area

The Amhara region is located in northwestern and north central Ethiopia and lies within 9°N and 13.45°N and 36°E and 40.30°E (Figure 1). The region has a total area of approximately 170,000 km² and is composed of eleven administrative zones. Elevation ranges from 506 to 4517 m above sea level. Annual rainfall varies geographically from about 770 mm in the lowlands to 2000 mm in the highlands. Mean annual temperature varies seasonally from 16°C in the rainy season to 27°C in the dry season. Rainfed agriculture is the main source of agricultural production and cropland and pasture are the major land cover types in the region [Lakew et al., 2000]. Malaria transmission in the region is unstable [Graves et al., 2009; Jima et al., 2010]. The major epidemic season is from September to December following the summer rainy season from June to August [Wimberly et al., 2012]. A previous study reported lagged positive associations of malaria with satellite-derived environmental factors such as rainfall, ET, and land surface temperature (LST) [Midekisa et al., 2012]. Large regional malaria epidemics occurred from 2003 to 2005 [FMOH, 2008; Jima et al., 2010; Wimberly et al., 2012]. Since that time, epidemics have generally been smaller and more localized, with some smaller outbreaks occurring in 2008 and 2009 [Midekisa et al., 2012; Wimberly et al., 2012]. Starting in

Figure 1. Location map of the Amhara region of Ethiopia showing regional variation in elevation from the Shuttle Radar Topography Mission (SRTM) topographic data along with the locations of sampling blocks and polygons.
2005–2007, the government of Ethiopia put in place an ambitious national strategy to control malaria; as a result, tens of millions of long-lasting insecticide-treated nets were distributed to all malarious areas all over the country including the Amhara region [Deressa et al., 2011; Jima et al., 2010].

*Anopheles arabiensis* is the principal malaria vector in Ethiopia [Habtewold et al., 2004; Taye et al., 2006; Tirados et al., 2006], including the Amhara region [Animut et al., 2013]. *Anopheles funestus* is also found in many parts of Ethiopia [Taye et al., 2006; Tirados et al., 2006]. The presence of *Anopheles gambiae* s.s. has been reported in the southern part of the country, but not in the Amhara region [Tirados et al., 2006]. The habitat preferences of *Anopheles arabiensis* are similar to those of *A. gambiae* s.s.; they both prefer sun lit and shallow freshwater pools, wells, and rice fields [Gimnig et al., 2001; Githeko et al., 1996; Shililu et al., 2003]. However, *A. arabiensis* is able to breed in a wider variety of habitats than *A. gambiae* s.s., including edges of receding streams, natural swamps, and other man-made habitats [Shililu et al., 2003]. *A. funestus* prefers larger water bodies with emergent vegetation [Gimnig et al., 2001; Walker and Lynch, 2007].

### 2.2. Wetland Mapping

#### 2.2.1. Satellite Data

Landsat scenes were obtained from the Landsat surface reflectance climate data records (CDR) data archive [Maiersperger et al., 2013; Masek et al., 2006], and included both Landsat 5 TM and Landsat 7 ETM+ data. These surface reflectance products were generated using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm by the U.S. Geological Survey. We used dry season Landsat acquisitions from October to May with less than 10% cloud cover and examined all the available data to confirm their suitability for mapping wetlands. We employed a local linear histogram matching technique to gap fill Landsat 7 ETM+ images with scan line corrector failure data using dry season Landsat 5 TM data [Storey et al., 2005]. The study area encompassed 11 Landsat scenes. We selected the best cloud-free image from 2008 to 2012 for each path-row combination and combined these scenes to create a regional mosaic.

We computed multiple optical spectral indices from the Landsat surface reflectance data. Normalized difference water index (NDWI) was derived using different band combinations such as bands 2 (0.52–0.60 μm) and 4 (0.77–0.90 μm) [McFeeters, 1996], bands 2 and 5 (1.55–1.75 μm) [Xu, 2006], and bands 4 and 5 [Gao, 1996]. Tasseled Cap (TC) indices that included TC brightness, TC greenness, and TC wetness were computed from Landsat surface reflectance bands using established methods [Huang et al., 2002]. Fractional vegetation cover (FVC), fractional soil cover (FSC), and fractional water cover (FWC) were derived as indicators of subpixel fraction of vegetation, soil, and water, respectively, based on multiple end-member spectral mixture analysis (MESMA). The MESMA model allows spectral end-members to vary between pixels and provides a subpixel fraction of spectra as a linear combination of end-members within a pixel [Powell et al., 2007; Rashed et al., 2003; Roberts et al., 1998]. We collected end-member spectra from Landsat imagery using manual interpretation and our knowledge of the landscape of the study site. The MESMA model produced a set of maps that represented a per-pixel fractional covers of end-member classes.

#### 2.2.2. Topographic Input Data

Because topography affects the spatial variability of soil moisture and surface water on the landscape, topographic indices are useful metrics to estimate these hydrological variations [Sorensen et al., 2006]. We used 30 m Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) data obtained through a special request to the National Geospatial Intelligence Agency, submitted through the U.S. Geological Survey. We used the DEM to calculate slope using Environmental Systems Research Institute (Esri) ArcGIS (10.1) software. In addition, topographic wetness index (TWI) was derived from the DEM data. We computed TWI, which provides an index of potential moisture availability, as the natural logarithm of the ratio of the upslope drainage area over slope [Sorensen et al., 2006].

#### 2.2.3. Training and Validation Reference Data

Wetlands and other land cover classes were identified from Google Earth archives that included SPOT and Digital Globe imagery for the 2008–2012 periods. We used a hybrid approach of manual digitizing using sampling blocks as well as semiautomated techniques using polygons to generate training and validation sample points. This approach was based on the methods used in other remote sensing studies that have used tree-based ensemble methods in combination with large training data sets to successfully map land cover across large regions [Bwangoy et al., 2010; Dong et al., 2012; Hansen et al., 2002, 2008]. First, a total of 105 gridded sampling blocks were selected from a 10 km × 10 km grid across the entire Amhara region.
We randomly split these 105 sample blocks into training (54 blocks) and validation (51 blocks). We then generated 12,699 training and 10,972 sample points in these sample blocks using a 350 m minimum linear distance among sampling points to help distribute the points evenly within the blocks. Each of the sample points was manually digitized to one of five land cover classes including open water, herbaceous wetland, woody vegetation, sparse vegetation, and cropland. We used photointerpretation techniques and our knowledge of the study area to assign these classes.

Second, we manually digitized 857 polygons of homogenous land cover classes. We randomly split these polygons into 433 training and 424 validation polygons (Figure 1). Our objective in selecting these polygons was to obtain additional training and validation data from geographic areas outside of the sampling blocks, and to increase our representation of wetlands and open water bodies, which were of primary interest for applications to mosquito-borne diseases. We randomly generated 42,718 training and 44,066 validation random sample points within these polygons. In total, we used 110,455 sample points, which included 55,417 training points and 55,038 validation points, derived from the combination of the sample blocks and polygons (example maps are provided in the supporting information, Figures S1 and S2). The overall proportion of the five land cover classes in the training points included open water (26%), herbaceous wetland (31%), woody vegetation (7%), sparse vegetation (7%), and cropland (29%), whereas the overall proportion in the validation points included open water (22%), herbaceous wetland (35%), woody vegetation (7%), sparse vegetation (10%), and cropland (26%).

2.2.4. Decision Tree Classification

Decision tree classification approaches have been widely used with remote sensing data for mapping land cover in general [DeFries et al., 1997; Hansen et al., 1996, 2008] and specifically for wetland mapping [Baker et al., 2007; Bwangoy et al., 2010; Corcoran et al., 2013; Tulbure and Broich, 2013; Wright and Gallant, 2007]. We employed a random forest decision tree classification algorithm to develop the spatially explicit land cover map, with a particular focus on identifying herbaceous wetlands, for the Amhara region of Ethiopia. Random forests are a nonparametric classification that consists of ensembles of decision trees [Breiman, 2001]. Decision trees use recursive partitioning to classify the dependent variable via a hierarchical set of binary splits of the predictor variables. The random forest algorithm creates multiple trees, in our case, 1000 trees for each classification. For each tree, a random subset of the observations is sampled with replacement. For each split within the tree, a subset of predictor variables is randomly chosen. The final random forests classification output for a given set of predictor variables is based on the majority vote of the classifications from all of the component trees. We carried out random forests modeling using the randomForest library in the R environment for statistical computing [R Core Team, 2014].

We evaluated the effectiveness of three models that used the following sets of predictor variables: (1) topographic indices from SRTM data, (2) individual reflectance bands and multispectral indices from Landsat TM/ETM+ imagery, and (3) combined spectral and topographic data. Overall percentage correct classification was computed for each model, and the user’s accuracy, producer’s accuracy, and specificity were computed for each of the land cover classes. User’s accuracy is the number of correctly classified validation points divided by the total number of points predicted to be in that class, whereas producer’s accuracy (identical to sensitivity) is the number of correctly classified validation points divided by the total number of validation points of that class. Specificity is the proportion of validation points correctly classified as not belonging to a class divided by the total number of points that do not belong to that class [Sader et al., 1995]. We also calculated the Kappa coefficient, which measures the agreement between observed and predicted classes by adjusting for the probability of agreement by chance [Cohen, 1968]. The Kappa coefficient ranges from 0 to 1, in which 1 represents prefect agreement and 0 represents no agreement. We used the independent validation data set to assess the performance of the three random forest models. A summary diagram of the data sets and methods used in the land cover classification is presented in Figure 2. We used 18 independent input variables in the classification algorithm for the best model (Table 1). A land cover map of the Amhara region of Ethiopia was produced for the 2010 year using the random forest decision tree classification algorithm based on the best model (Figures 3 and 4).

2.3. Spatial Association of Wetlands and Malaria

We obtained district-level monthly malaria case data for the periods 2007–2009 through the Health, Development and Anti-Malaria Association (HDAMA) in Addis Ababa, Ethiopia. Historical surveillance data from
the integrated disease surveillance and response (IDSR) system were collected [Jima et al., 2012]. These data included monthly reports of clinically diagnosed outpatient cases and total number of outpatient visits [Wimberly et al., 2012]. The district-level proportion of outpatients clinically diagnosed with malaria during the major epidemic season (September to December) from 2007 to 2009 was used as an indicator of the spatial variability in malaria risk. We employed the proportion of outpatient cases because it controls for variation across districts in the numbers of individuals accessing the health system [WHO, 2006].

We used the 2010 land cover map to compute percent area of the herbaceous wetland class for each district (Figure 5). The total number of malaria cases was averaged over 2007–2009 and summarized for the three seasons, including the dry (January to April), wet (May to August), and fall (September to December) seasons (Figure 5). We selected 38 districts for the dry season, 37 districts for the wet season, and 39 districts for the fall season based on availability of both malaria and total outpatient cases for the 2007–2009 period. The summaries of climate variables and malaria in the three seasons are shown in Table 2. The total number of malaria cases was the highest in the fall season and lowest in the dry season. The wettest and coldest climate conditions were observed in the wet season, whereas the driest and warmest climate conditions occurred during the dry season.

We examined the spatial association of percent of wetlands with the proportion of outpatients clinically diagnosed with malaria using negative binomial generalized linear models (GLM). The total number of outpatients clinically diagnosed with malaria was the dependent variable, and the natural logarithm of the total number of outpatients was used as an offset. We used the average of satellite-derived climatic variables

![Figure 2. Summary diagram of input data and methods used in the land cover classification.](image)

![Table 1. Variable Importance of the Random Forests Model, Expressed as the Percent Reduction in Prediction Accuracy When the Variable Is Omitted From the Model (Model 3).](table)
for the 3 years preceding each season. These climatic variables were summarized over May 2006 to April 2009 for the dry (January to April) season models, September 2006 to August 2009 for the wet (May to August) season models, and January 2007 to December 2009 for the fall (September to December) season models. These data included the 8 day MODIS Terra LST (MOD11A2) product at a 1 km spatial resolution and the Tropical Rainfall Measuring Mission (TRMM) product (3B42) to estimate rainfall. Annual averages of mean LST and total rainfall were calculated for each district and used as independent variables. The district-level percent of wetlands from our land cover map and mean SRTM-based elevation were also used as independent variables. Because the percent of wetlands variable was skewed, we used a square root transformation to reduce the influences of extreme values and meet the statistical assumptions of the GLM model. We conducted multicollinearity tests using variance inflation factor (VIF) analysis, and found that all variables had a VIF less than 2, which indicated that correlation among the explanatory variables was not an issue in the GLM model. We ran two types of negative binomial GLM models: model 1 (without wetlands) and model 2 (with wetlands) to test for spatial association of wetlands with malaria while controlling for the effects of temperature, precipitation, and elevation.

3. Results

The random forest classification for model 3 produced the most accurate wetland maps with an overall accuracy of 93.5% with a Kappa coefficient of 0.91 while model 2 (only spectral data) and model 1 (only topographic data) yielded lower overall accuracies (88.3% and 80.2%) and Kappa coefficients (0.84 and 0.73), respectively. In addition, validation results indicated that herbaceous wetland class showed higher

Figure 3. Land cover classification from model 3 (spectral and topographic variables) for the Amhara region of Ethiopia in 2010. Subset locations a, b, c, and d are displayed in Figure 4
producer’s and user’s accuracies for model 3 (93.5% and 95.3%) as compared to model 2 (92.2% and 84.2%) and model 1 (82.5% and 79.45%), respectively (Figure 6 and Table 3). Producer’s accuracy provides an estimate of omission error while user’s accuracy measures commission error [Sader et al., 1995]. The herbaceous wetland class also had higher specificity for model 3 (96.2%) as compared to model 2 (90.1%) and model 1 (87.1%) (Figure 6 and Table 3). Further, visual comparisons of the land cover classification with both Landsat

![Figure 4. Comparison of land cover classification with imagery: (left) Landsat band composites of 543 for the year 2010; (middle) modeled land cover classification for the year 2010, and (right) high-resolution imagery from Google Earth. Sites a, b, c, and d are referenced in Figure 3.](image-url)
and higher-resolution imagery from Google Earth showed a reasonably accurate spatial depiction of the land cover classes, particularly herbaceous wetlands, in the study area (Figure 4).

These results confirmed that using a multisensor (Landsat TM/ETM+ and SRTM) remote sensing approach to mapping land cover classes in general and wetlands in particular can improve the accuracy of classification. The independent input variables used in the best model (model 3) are listed in descending order of the decrease in prediction accuracy that resulted when the variable was omitted from the model (Table 1).

The top five most important variables in the best classification model (model 3) included a combination of topographic indices (e.g., elevation, slope, and topographic wetness index), and spectral variables (Landsat band 3, and fractional water cover) as shown in Table 1. These findings showed that both spectral and topographical input data were important for the land cover classification. However, it should be noted that the importance values of the individual spectral indices were likely reduced because there was a relatively large number of correlated spectral variables. Because there were fewer individual topographic variables and they were less correlated with the other predictor variables, they tended to have higher importance values even though the topographic model (model 1) had the lowest accuracy statistics.

Results from the generalized linear model based on the proportion of outpatient malaria cases showed that there was a significant positive spatial association of percent wetlands and malaria incidence for the fall malaria season \( (p \text{ value } = 0.001) \), for the wet malaria season \( (p \text{ value } = 0.001) \), and for the dry season \( (p \text{ value } = 0.03) \) (Table 4). Percent wetlands had a positive relationship with the number of outpatient
malaria cases after controlling for climatic variability and correcting for spatial variability in the numbers of outpatients seeking care (scatterplots of malaria indicators versus percent wetlands are provided in the supporting information, Figure S3). We used analysis of variance (ANOVA) based on chi-square statistics to compare the full model (with wetlands) with a reduced model (without wetlands) for each malaria transmission season. These results showed that, for all seasons, the addition of percent of wetland in the full model significantly improved model fit as compared to reduced model as indicated by the chi-square statistics for the fall malaria season (p value = 0.004 with reduced model degree of freedom (DF) = 35 and full model DF = 34), the wet malaria season (p value = 0.004 with reduced model DF = 32 and full model DF = 31), and the dry malaria season (p value = 0.04 with reduced model DF = 34 and full model DF = 33).

4. Discussion

4.1. Wetland Classification

Tropical wetlands provide various beneficial ecosystem services that include conservation of biodiversity, regulation of the hydrological cycle, sequestration of carbon, and improvement of water quality [Bwangoy

| Season | Malaria Cases Mean (SD) | Malaria Proportion* Mean (SD) | Rainfall (mm) Mean (SD) | LST (°C) Mean (SD) |
|--------|-------------------------|-------------------------------|-------------------------|---------------------|
| Dry (n = 38) | 2288 (2399) | 19.62 (13.66) | 97.00 (22) | 25.52 (2.29) |
| Wet (n = 37) | 2536 (2846) | 22.36 (16.83) | 786.32 (139.93) | 20.85 (2.96) |
| Fall (n = 39) | 3882 (4955) | 25.98 (18.23) | 152.32 (50.34) | 22.22 (2.47) |

*Malaria Proportion = the percentage of outpatient malaria cases among total outpatient cases.

Table 2. Descriptive Statistics of Climate Variables and Malaria in the Three Seasons Including Dry (January to April), Wet (May to August), and Fall (September to December) Over the 2007–2009 Period

Figure 6. Producer’s (top) accuracy, (middle) specificity, and (bottom) user’s accuracy across the three models: model 1 (with topographic data), model 2 (with spectral data), and model 3 (with combined topographic and spectral data). The abbreviations in the x axis include Sparse = sparse vegetation, Water = open water, Wetland = herbaceous wetland, and Woody = woody vegetation.
et al., 2010; Li and Chen, 2005; Segers, 1998; Teferi et al., 2010; Tiner, 2003]. However, wetlands can also facilitate the transmission of mosquito-borne diseases, including malaria, by serving as potential breeding habitats for the vector mosquitoes. In Ethiopia, wetland degradation has been reported due to various factors such as agricultural encroachment, livestock grazing, population pressure, and infrastructure construction [Abebe and Geheb, 2003]. Thus, there is a need for an integrated management strategy that maximizes useful ecosystem services and minimizes the human health impacts of wetlands. Accurate mapping of wetlands is a crucial first step toward this goal.

The study presented here demonstrated the utility of multisource remote sensing data to characterize the spatial distribution of herbaceous wetlands in the context of malaria transmission in the Amhara region of Ethiopia. This finding was based on the evaluation of the three classification models using high spatial resolution reference imagery from Google Earth. Results showed that model 3 based on combined spectral data and topographic data yielded the highest overall accuracy. This finding is consistent with previous research that has demonstrated the importance of employing multisource remote sensing data, compared to single sources, to improve wetland classification [Baker et al., 2007; Bwangoy et al., 2010; Corcoran et al., 2013; Wright and Gallant, 2007]. The spatial distribution of herbaceous wetlands in the study region is constrained by local topography and characterized by higher soil moisture. Therefore, wetland mapping was enhanced by utilizing spectral information in the optical and infrared bands to characterize vegetation and surface moisture along with topographic indices computed using a digital elevation model that was derived from synthetic aperture radar data.

We found a producer’s accuracy of 95.3% and a user’s accuracy of 93.5% for the wetland class, in concordance with previous studies that mapped wetlands and reported overall accuracy greater than 90% in different regions [Bwangoy et al., 2010; Tulbure and Broich, 2013; Wright and Gallant, 2007]. A previous study in the Choke mountain range, a subcatchment within our study area, also characterized wetlands using Landsat data to investigate changes in land cover from 1986 to 2005; their wetland classification results yielded overall accuracies of 94.1% and 93.5% for 1986 and 2005, respectively [Teferi et al., 2010]. However, their study did not include a detailed examination of the relative contribution of spectral bands and indices used in the classification. One of the challenges of land cover classification at the current Landsat scale (30 m) is to represent land cover features when multiple classes (e.g., vegetation and water) are present within a pixel. Earlier research has demonstrated the benefits of using subpixel representation of spectra for land cover mapping [Powell et al., 2007; Rashed et al., 2003; Roberts et al., 1998], but there has been less effort to utilize the approach for wetland mapping. To address this problem, this study employed a subpixel representation of spectra based on fractional cover for different land cover classes within a pixel. Our findings showed that the addition of subpixel level fractional cover spectral data improved the classification results; in particular, fractional water cover was one of the top five most important variables in the best model. The implication is that Landsat-scale subpixel representation of land surface spectra could be employed to improve wetland identification in other geographic settings.

### 4.2. Spatial Association of Wetlands and Malaria

Spatial heterogeneity in malaria epidemiology is influenced by environmental factors such as distance from the nearest mosquito breeding site and availability of permanent and temporary water bodies [Bousema et al., 2010; Kreuels et al., 2008]. For example, herbaceous wetlands provide moisture needed to complete the aquatic developmental stages of mosquitoes. Our findings showed statistically significant positive

---

**Table 3. Accuracy Assessment of the Three Land Cover Classification Models**

| Model | Metrics     | Cropland | Sparse Vegetation | Open Water | Herbaceous Wetland | Woody Vegetation |
|-------|-------------|----------|-------------------|------------|--------------------|------------------|
|       | Producer’s | 68.9     | 80.7              | 95.4       | 82.5               | 66.9             |
|       | Specificity| 88.9     | 97.9              | 98.9       | 87.1               | 97.1             |
|       | User’s     | 69.2     | 84.1              | 96.9       | 79.5               | 68.4             |
| Model 2 | Producer’s | 84.4     | 72.9              | 96.4       | 92.2               | 75.7             |
|       | Specificity| 94.8     | 98.4              | 99.9       | 90.1               | 99.4             |
|       | User’s     | 85.6     | 84.6              | 99.67      | 84.2               | 92.2             |
| Model 3 | Producer’s | 91.0     | 87.8              | 99.0       | 95.3               | 85.0             |
|       | Specificity| 96.2     | 99.0              | 99.9       | 96.2               | 99.4             |
|       | User’s     | 89.3     | 91.3              | 99.7       | 93.5               | 92.4             |
spatial association of percent of wetlands and malaria incidence at the district level during the dry, wet, and fall malaria transmission seasons. Earlier studies also provided evidence that anopheline mosquitoes are associated with wetland habitats. A study in the Tigray region of Ethiopia reported the highest percentage of larval samples of *Anopheles arabiensis* in natural wetlands such as swamps and marshes [Yohannes et al., 2005]. Moreover, anopheline mosquito larvae samples were abundantly found in the edges of rivers and natural swamps in a study in the Rift Valley region of Ethiopia [Kenea et al., 2011]. In western Kenya, research results showed that *A. gambiae s.s.* preferred wetter soil for oviposition as a mechanism of surviving during the moisture limited dry season [Minakawa et al., 2001]. A study in the Kilombero valley of East Africa showed that there was increased abundance of *Anopheles arabiensis* in villages closer to river valleys [Charlwood et al., 2000].

Previous studies have examined the spatial relationships of wetlands and malaria in relatively small landscapes. For example, a recent study in Zambia reported that malaria incidence followed seasonal patterns such that lower transmission in the dry season was confined to the floodplain while higher transmission expanded to the broader region in the wet season over 5 years of the study period [Shiff et al., 2013]. Their study emphasized the potential role of wetlands as hot spots for malaria transmission and their interactions with seasonality. However, knowledge of the spatial association of malaria and wetlands is lacking at broader regional scales, particularly for the Amhara region of Ethiopia. Our study is one of the first to assess the spatial association of malaria and wetlands across a broader region. The results showed that the spatial association of malaria and wetlands were statistically significant during all parts of the year, including the dry season when precipitation is the major limiting factor for mosquito development, the wet season when temperature is the major limiting factor, and the fall season when large malaria epidemics historically occurred. This information can assist public health decision making in efforts to control and eliminate malaria because it provides important information on the location of potential high-risk areas for malaria transmission.

Wetlands likely have the largest direct effect on malaria transmission in populations living nearby [Ernst et al., 2006]. For example, higher malaria risk was reported in households living closer to mosquito breeding sites in the Oromia region of Ethiopia [Peterson et al., 2009]. However, wetlands may also have broader influences on seasonal patterns of malaria transmission by serving as focal areas that sustain transmission during the dry season when mosquito abundance and parasite infection rates in humans are low [Bousema et al., 2012]. As a result, malaria transmission maintained throughout the dry season can then be a possible starting point for initiating epidemics in the high transmission season. Identification of these wetland habitats can thus be a key step toward successful intervention strategies for malaria control and elimination [Kelly et al., 2012; Moonen et al., 2010]. The use of multisource remote sensing data to quantify the spatial distribution of wetlands may also make an important contribution to malaria early warning systems [Abeke et al., 2004; Thomson and Connor, 2001] by identifying areas likely to have distinctive seasonal patterns of malaria transmission and persistent high levels of malaria risk.

In summary, the information gained in this study about the spatial distribution of herbaceous wetlands in the context of malaria transmission can help to support targeted interventions to effectively control and
eliminate malaria in the region. The results provide multiple sources of evidence supporting the spatial association of malaria and wetlands based on multisensor remote sensing and epidemiological surveillance data. However, a limitation of this study is that it employed district-level passive surveillance data to analyze environmental risk factors and malaria cases; as a result, finer-scale variation associated with the distribution of wetlands within subdistricts level cannot be addressed. Future research could use our wetland maps as a starting point to conduct active surveillance at a household level to advance our understanding of the dynamic relationship between malaria and wetlands at a finer spatial scale. We did not assess the impact of other nonenvironmental factors that can influence the spatial heterogeneity of malaria epidemiology, including public health interventions such as indoor residual spraying and mass distribution of long-lived insecticide-treated nets, socioeconomic status, human population movements, and levels of immunity in the human population [Breman, 2001; Clark et al., 2008]. Although clinically diagnosed cases are the most available surveillance data in the study region, they can underestimate the prevalence of malaria; thus, future research should also address asymptomatic malaria cases [Breman, 2001]. Longer-term entomological studies are also needed to test the hypothesis that wetlands serve as refugia for mosquitoes and malaria transmission during the dry season.

5. Conclusions

In this study, we demonstrated the utility of multisensor remote sensing data to identify the spatial distribution of herbaceous wetlands for public health applications. We developed the most extensive map of wetlands currently available for the Amhara region of Ethiopia; this information can help to detect foci of malaria transmission. Our results showed that there was strong spatial association of malaria cases and wetlands at the district level across the dry, wet, and fall seasons. This finding emphasizes that efforts aimed at analyzing and mapping regional patterns of malaria risk should consider the influences of physiographic patterns and land cover characteristics in addition to climatic gradients. Wetlands can influence malaria in nearby areas by providing larval habitat for vector mosquito breeding, and may help to sustain mosquito populations and low levels of malaria transmission during the dry season. The land cover map developed in this study can thus help public health personnel to target intervention in control and elimination strategies in the Amhara region and can be used to support more detailed, finer-scale studies in the future to explore the mechanisms that link wetlands and malaria transmission.

Acknowledgments

This work was supported by NASA, grant NNX11AP79H, through the NASA Earth and Space Science Fellowship (NESSF) program and by National Institute of Allergy and Infectious Diseases through grant R01AI079411. We acknowledge the Amhara National Regional State Health Bureau for their cooperation. Abeke Mihretie, Paulos Semuniguse, and Hiwot Teka from the Health, Development, and Anti-Malaria Association collected the malaria surveillance data. Landsat surface reflectance products are courtesy of the U.S. Geological Survey (USGS). The MODIS data were obtained from the NASA Land Processes Distributed Active Archive Center (LP DAAC), located at the USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD, USA (http://lpdaac.usgs.gov). We would like to thank Jeffrey J. Danielsen at the USGS/EROS Center, Sioux Falls, SD, USA for providing the DEM data and Nicholas Humphrey at South Dakota State University for digitizing Google Earth reference data. Data are available on request from the corresponding author.

References

Abbe, D. Y., and K. Geheb (2003), Wetlands of Ethiopia, paper presented at Seminar on the Resources and Status of Ethiopia’s Wetlands, Int. Union for Conserv. of Nat. and Nat. Resour., Addis Ababa.

Abeku, T. A., S. I. Hay, S. Ochola, P. Langi, B. Beard, S. J. de Vlas, and J. Cox (2004), Malaria epidemic early warning and detection in African highlands, Trends Parasitol., 20(9), 400–405.

Aniunor, A., M. Balkew, and B. Lindtjorn (2013), Impact of housing condition on indoor-biting and indoor-resting Anopheles arabiensis density in a highland area, central Ethiopia, Malaria J., 12, 393, doi:10.1186/1475-2875-12-393.

Baker, C., R. L. Lawrence, C. Montagne, and D. Patten (2007), Change detection of wetland ecosystems using Landsat imagery and change vector analysis, Wetlands, 27(3), 610–619.

Beck, L. R., et al. (1994), Remote-sensing as a landscape epidemiologic tool to identify villages at high-risk for malaria transmission, Am. J. Trop. Medicine Hyg., 51(3), 271–280.

Bouyssou, T., et al. (2010), Identification of hot spots of malaria transmission for targeted malaria control, J. Infectious Diseases, 201(11), 1764–1774, doi:10.1086/652456.

Bouyssou, T., J. T. Griffin, R. W. Sauerwein, D. L. Smith, T. S. Churcher, W. Takken, A. Ghani, C. Drakeley, and R. Gosling (2012), Hitting hot spots: Spatial targeting of malaria for control and elimination, PLoS Medicine, 9(1), e1001165, doi:10.1371/journal.pmed.1001165.

Breman, L. (2001), Random forests, Mach. Learning, 45(1), 5–32.

Breman, J. G. (2001), The ears of the hippopotamus: Manifestations, determinants, and estimates of the malaria burden, Am. J. Trop. Medicine Hyg., 64(1-2), 1–11.

Brown, H. E., M. A. Diuk-Wasser, Y. Guan, S. Caskey, and D. Fish (2008), Comparison of three satellite sensors at three spatial scales to predict larval mosquito presence in Connecticut wetlands, Remote Sens. Environ., 112(23), 2301–2308.

Chabot-Couture, G., K. Nigmatulina, and P. Eckhoff (2014), An environmental data set for vector-borne disease modeling and epidemiology, PLoS ONE, 9(7), e103922, doi:10.1371/journal.pone.0103922.

Chase, J. M., and T. M. Knight (2003), Drought-induced mosquito outbreaks in wetlands, Ecol. Lett., 6(11), 1017–1024, doi:10.1046/j.1461-0248.2003.00533.x.

Chuang, T. W., and M. C. Wimerberly (2012), Remote sensing of climatic anomalies and West Nile virus incidence in the Northern Great Plains of the United States, PLoS ONE, 7(10), e46882, doi:10.1371/journal.pone.0046882.
Machault, V., C. Vignolles, F. Pages, L. Gadiaga, A. Gaye, C. Sokhna, J. F. Trape, J. P. Lacaux, and C. Rogier (2010), Spatial heterogeneity and temporal evolution of malaria transmission risk in Dakar, Senegal, according to remotely sensed environmental data, *Malaria J.*, 9, 252, doi:10.1186/1475-2875-9-252.

Machault, V., C. Vignolles, F. Borchi, P. Younatso, F. Pages, S. Birolant, J. P. Lacaux, and C. Rogier (2011), The use of remotely sensed environmental data in the study of malaria, *Geospatial Health*, 5(2), 151–168.

Maers sperger, T. K., P. L. Scaramuzza, L. Leigh, S. Shrestha, K. P. Gallo, C. B. Jenkerson, and J. L. Dwyer (2013), Characterizing LEDAPS surface reflectance products by comparisons with AERONET, field spectrometer, and MODIS data, *Remote Sens. Environ.*, 136, 1–13.

Martens, P., R. S. Kovats, S. Nijhof, F. de Vries, M. T. J. Livermore, D. J. Bradley, J. Cox, and A. J. McMichael (1999), Climate change and future populations at risk of malaria, *Global Environ. Chang.*, 9, S89–S107, doi:10.1016/S0959-3780(99)00020-5.

Martens, W. J. M., L. W. Niessen, J. Rotmans, T. H. Jetten, and A. J. McMichael (1995), Potential impact of global climate-change on malaria risk, *Environ. Health Perspect.*, 103(5), 458–464.

Masek, J. G., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, F. Gao, J. Kutler, and T. K. Lim (2006), A Landsat surface reflectance dataset for North America, 1990–2000, *IEEE Geosci. Remote. Sci.*, 3(1), 68–72, doi:10.1109/LGRS.2005.857030.

McFeeters, S. K. (1996), The use of the normalized difference water index (NDWI) in the delineation of open water features, *Remote Sens. Environ.*, 56(1–2), 241–252.

McRoberts, N., J. Lopez, J. D. Ninagawa, P. G. Tornero, N. L. Huertas, and J. J. Cupo (2010), Anopeline mosquito survival strategies during the dry period in western Kenya, *Biogeochemistry*, 97, 267–279, doi:10.1007/s10533-009-9370-8.

Mishnezimana, E., et al. (2005), Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands, *J. Med. Entomol.*, 42(3), 232–239.

Mitsch, W. J., and J. G. Gosselink (2007), *Ecology and Management of Wetlands*, 3rd ed., John Wiley & Sons, New York.

Moe, E., and E. Moe (2013), The role of spatial synchrony in malaria transmission, *Proc. Natl. Acad. Sci. U. S. A.*, 110(1), 388–392, doi:10.1073/pnas.0901356106.

Mugyensi, J., G. M. Henebry, and T. W. Chuang (2011), Remote sensing for early warning in the highlands of Ethiopia, *Malar. J.*, 10, 151, doi:10.1186/1475-2875-10-151.

Musenze, E., et al. (2013), Remote sensing for early warning in the highlands of Ethiopia: A multilevel analysis, *Malar. J.*, 12, 157, doi:10.1186/1475-2875-12-157.

Njuguna, J. G., and J. C. Beier (2013), Remote sensing for early warning in the highlands of Ethiopia, *Malar. J.*, 12, 191, doi:10.1186/1475-2875-12-191.

Njenga, S. K., et al. (2012), Remote sensing for early warning in the highlands of Ethiopia, *Malar. J.*, 11, 165, doi:10.1186/1475-2875-11-165.

Oromie, L., and J. C. Beier (2013), Remote sensing for early warning in the highlands of Ethiopia, *Malar. J.*, 12, 157, doi:10.1186/1475-2875-12-157.

Parsons, J. E., M. A. Midekisa, M. Pettersson, R. Wolfe, and D. O. Sembodja (2013), Remote sensing for early warning in the highlands of Ethiopia, *Malar. J.*, 12, 157, doi:10.1186/1475-2875-12-157.

Pepin, J. M., C. Vignolles, S. Birolant, J. L. Dwyer, and W. T. Wu (2008), Remote sensing for early warning in the highlands of Ethiopia, *Malar. J.*, 7, 168, doi:10.1186/1475-2875-7-168.

Peterson, L. N., B. H. Anderson, F. P. de Vries, and M. J. V. Jones (2010), Remote sensing for early warning in the highlands of Ethiopia, *Malar. J.*, 9, 126, doi:10.1186/1475-2875-9-126.

Phiri, M., and J. C. Beier (2013), Remote sensing for early warning in the highlands of Ethiopia, *Malar. J.*, 12, 157, doi:10.1186/1475-2875-12-157
Wright, C., and A. Gallant (2007), Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data, Remote Sens. Environ., 107(4), 582–605, doi:10.1016/j.rse.2006.10.019.

Xu, H. Q. (2006), Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery, Int. J. Remote Sens., 27(14), 3025–3033, doi:10.1080/01431160600589179.

Yohannes, M., M. Haile, T. A. Ghebreyesus, K. H. Witten, A. Getachew, P. Byass, and S. W. Lindsay (2005), Can source reduction of mosquito larval habitat reduce malaria transmission in Tigray, Ethiopia?, Trop. Medicine Int. Health, 10(12), 1274–1285.

Zhang, Y., P. Bi, and J. E. Hiller (2008), Climate change and the transmission of vector-borne diseases: A review, Asia Pac. J. Public Health, 20(1), 64–76, doi:10.1177/1010539507308385.