A Taxonomy of Unmixing Algorithms Using Li-Strahler Geometric-Optical Model and other Spectral Endmember Extraction Techniques for Decomposing a QuickBird Visible and Near Infra-Red Pixel of an *Anopheles arabiensis* Habitat

Benjamin G. Jacob*[,1], Joseph M. Mwangangi[2], Charles M. Mbogo[2] and Robert J. Novak[1]

[1]*School of Medicine, Division of Infectious Diseases, University of Alabama at Birmingham, 845 19th Street South, Birmingham, AL 35294-2170, USA;* [2]*Kenya Medical Research Institute (KEMRI), Centre for Geographic Medicine Coast, PO Box 230, Kilifi, Kenya*

Abstract: Spectral unmixing algorithms have proliferated in a variety of ecological disciplines by exploiting remotely-sensed data. However, in East African rice field agro-ecosystems, aquatic habitats of *Anopheles arabiensis,* a major vector of malaria in Sub-Saharan Africa (SSA), utilize semi-permanent to temporary habitats (e.g., floodwater areas, vernal pools, hoof prints) which pose a special problem in landscape studies, basically one of spatial scale. For example, low spatial resolution pixel sizes from satellite sensors are often too large for identification of productive riceland *An. arabiensis* habitats. In this research we spectrally decomposed a sub-meter spatial resolution (i.e., QuickBird) riceland *An. arabiensis* habitat pixel for predicting productive habitats in a riceland environment. Initially, we constructed a regression model which revealed that paddy preparation *An. arabiensis* habitats were the most productive based on spatiotemporal field-sampled count data. Individual pixel spectral reflectance estimates from a QuickBird visible and near-infra-red (NIR) at 0.61m spatial resolution data of a paddy preparation *An. arabiensis* habitat were then extracted by using a Li-Strahler geometric-optical model. The model used three scene components: sunlit canopy (C), sunlit background (G) and shadow (T) generated from the riceland image. The G, C, T components' classes were estimated using ENVI, an object-based classification algorithm. In ENVI®, the Digital Number (DN) of the pixel in every QuickBird band was viewed using the *z*-profile from a spectral library. After making an atmospheric correction from the image for the study site, the DN was converted into ground reflectance. A convex geometrical model was also used for endmember validation of the spectrally decomposed paddy preparation habitat. An ordinary kriged-based interpolation was performed in ArcGIS® Geostatistical Analyst using the reference signature generated from the unmixing models. Linear unbiased predictors and variance estimates were derived of all productive *An. arabiensis* habitats in the study site based on the extracted pixel endmember reflectance estimates. Spectral unmixing tools may be used to decompose QuickBird visible and NIR pixel reflectance of a productive *An. arabiensis* habitat. Thereafter, an ordinary interpolator can use the sub-pixel data along with other spatially continuous explanatory variables sampled from productive habitats for targeting other high density foci habitat sites which can help implement larval control strategies in a riceland environment.

Keywords: QuickBird, *Anopheles arabiensis*, endmember spectra, Li-Strahler geometric –optical.

1. INTRODUCTION

Homogenous habitat pixels of *An. arabiensis*, a major vector of malaria in Sub-Saharan Africa (SSA) can be extracted from satellite data using object based classification (e.g., ENVI® technology) and then analyzed using traditional classification algorithms (e.g., nearest-neighbor, minimum distance, maximum likelihood), or knowledge-based approaches and fuzzy classification [1]. Generally, the algorithms behind image classification methods utilize spectral, spatial, texture, shape, context and ancillary information to model spatial objects and land use land cover (LULC) attributes based on individual pixel information [2]. In the preprocessing of optical bands it is then possible to calculate both the reflectance as it is measured by the sensor and the reflectance coming from the pixel at sensor level [3]. Additionally, local variability in remotely sensed data can be characterized by computing statistics from *An. arabiensis* habitat pixel spectral endmember data, e.g., coefficient of variance or autocovariance, or by analysis of fractal relationships [1].

Recently, there have even been some attempts to improve the spectral analysis of remotely sensed environmental data by using texture transforms in which some measure of variability in pixel values [i.e., digital number (DN)] is estimated within local windows; e.g. contrast between neighboring pixels [4]; the standard deviation [5], or local variance [6]. The coefficient of variance gives a measure of the total relative variation of pixel values by providing a normalized measure of dispersion of a probability distribution using the ratio of the standard deviation to the

*Address correspondence to this author at the School of Medicine, Division of Infectious Diseases, University of Alabama at Birmingham, 845 19th Street South, Birmingham, AL 35294-2170, USA; Tel: (205) 966-7894; Fax: (205)934-5600; E-mail: bjacob@uab.edu

1875-4139/11

mean [2]. Previous programs have been presented to compute spatial co-occurrence matrices from digital data in image format for performing spectral texture analysis for pixel based analyses but were restricted to physical characteristics of terrain-related parameters, primarily derived from digital elevation models. Covariates of *An. arabiensis* habitats, however, include multiple field and remote-sampled predictor variables [1]. Furthermore, most classical mathematical algorithms for spectral image classification do not usually quantify the dependence existing between a pixel and its neighbors, i.e., spatial autocorrelation. Autocorrelation is a characteristic of data derived from a process that is articulated in one or more spatial dimensions which can describe the error structure of ecological sampled data. Since the present methodologies for image classification can allow the multispectral satellite signal with an LULC to vary from pixel to pixel about a mean spectrum, variance covariance matrices can quantify the variability around the mean. For example, an autocorrelation consistent covariance matrix using generalized methods of moments (GMM) may be used in reparameterization and parsimonious modeling of sub-pixel endmemeber spectra [2]. Furthermore, by adding radiometric band and sensor complementary information related to the textural features of an image, the autocorrelation spatial structure of the DN can also be further quantified [5]. In this way, the results obtained from pixel-by-pixel classifiers can simultaneously take into account both radiometric and texture information in *An. arabiensis* habitat surface reflectance components. This improvement would arise from the hypothesis that a pixel is not independent of its neighbors and, furthermore, that its dependence can be spatially quantified and incorporated into the classifier.

Introducing geostatistical texture into an *An. arabiensis* habitat classification process using a set of measures of spatial autocorrelation indices based on univariate and multivariate estimators of a variogram function may also enable forecasting prolific habitat sites. For example, the autocorrelation structure of a decomposed *An. arabiensis* habitat endmember spectra can be modeled by fitting a theoretical variogram model to the empirical variogram using the following parameters: (i) model family, such as exponential, spherical, or Gaussian; (ii) nugget variance, (i.e., the variance among adjacent samples); (iii) range, or the distance beyond which observations are spatially independent; and (iv) sill, the constant variance among spatially uncorrelated samples. Predictors of *An. arabiensis* habitats can then be generated by incorporating a model of the covariance using the random function on a weighted moving average interpolation. Thus, extracting individual constituent spectra and a set of corresponding fraction abundances present in a pixel from a productive *An. arabiensis* habitat based on spatiotemporal field-sampled count data can generate a spectral 'footprint' which may be used for discriminating other productive habitats throughout the rice cycle. Treatments or habitat perturbations should be based on surveillance of larvae in the most productive areas of an ecosystem [1].

Traditionally, pixel-dependent classification processes assign labels to individual pixels [2] which can then be used to predict single pixels extracted from satellite data or the proportional membership of each pixel to a specific class [7]. The per-pixel classification includes the hard classification [e.g., Maximum likelihood classification(MLC) [8, 9], ISODATA classification [10, 11], and the soft classification (e.g. linear spectral unmixing [12-14] and neural network [15, 16]. Of these pixel classification techniques the soft classification provides the most realistic representation of land cover. However, the per-pixel hard classification cannot by itself utilize individual pixel reflectance estimates for an *An. arabiensis* habitat classification due to the spectral error caused by the effect of the excess heterogeneous surface topographic-induced illumination variants in riceland agroecosystems, which can result in extensive pixel misclassification. Land cover change throughout the rice cycle (i.e., flooding, usage of traditional and power tillers, low-lift irrigation pumps, pesticides on selected land and soil qualities) can lead to a mixed spectral *An. arabiensis* habitat pixels [17].

Commonly, mixed pixels occurs as a result of each sensor element's instantaneous field-of-view (iFOV) imaging more than one land cover class [18], which often results in poor classification accuracy especially when conventional algorithms such as the MLC are used [19]. The iFOV is the angle subtended by the geometrical projection of single detector element to the Earth's surface [7]. A mixed pixel is a combination of spectra from multiple unique substances (i.e., endmembers) [3]. Endmembers for riceland *An. arabiensis* habitats commonly represent disparate macroscopic objects (e.g., mud, rice plant leaves) [1]. Mixed pixels will complicate direct quantitative interpretation of an *An. arabiensis* habitat as the upwelling radiance from multiple materials will be integrated into a single observed spectrum. The combination of finite pixels and heterogeneous riceland landscapes will result in excessive spectral mixing within an *An. arabiensis* habitat pixel.

In order to resolve the mixed pixel problem different models have been developed to unmix the pixels into different proportions of their endmembers using various combinations of spectral unmixing algorithms and geometric-optical models. Pixel unmixing is the decomposition of mixed pixels into a collection of distinct endmembers and a set of fractional abundances that indicate the proportion of each endmember in the mixed pixel [20]. The approaches vary in their methods, and through their algorithmic formulations implicitly incorporating assumptions regarding the physical mechanisms and mathematical structure by which the reflectance properties from disparate substances combine to yield mixed pixel spectra. Since determination of endmembers would be the fundamental stage in each process of an *An. arabiensis* habitat classification, selection of an unmixing algorithm is vital.

A linear mixing model may be used to separate endmembers of a mixed pixel from a productive *An. arabiensis* habitat based on field-sampled count data, which,

in turn, may be used to forecast other high density foci habitats in riceland agro-ecosystems. Linear mixed models have been applied in many fields including air quality, chromatography, and spectroscopy and more recently in hydrochemical studies of natural catchments. A linear unmixing algorithm can predict retrieved fractional abundances and their associated errors due to both natural variability and corrupting noise sources [21]. The linear mixing model assumes that the mixed pixel is a linear combination of ground cover radiance spectra. One of the most common approaches in linear mixed modeling is an iterative generalized least squares estimation (LSE) procedure [22, 23]. The LSE can generate individual unbiased values with a minimum variance of prediction error provided that selection is within levels of fixed effects and the variances of the random effects before selection are known [23]. Recently, the combination of linear spectral unmixing and three-dimensional discrete wavelet transforms (DWT) for pixel feature extraction has been developed and refined [24, 25]. In particular, one of the refinements has been in the fast implementation of the three-dimensional wavelet transform, denoted as the hyperspectral DWT using the lifting implementation of the Haar wavelet basis in all three dimensions [26, 27]. The other refinement has been in replacement of least squares matrix pseudo inversion in the linear spectral unmixing technique with more numerically stable singular value decompositions (SVD) [28]. The SVD allows for a figure of merit calculation on the effectiveness of the endmember spectral signature library used in unmixing. Object variations that result in pixel variations within spectral signatures for surface material require various unmixing algorithms using multitemporal remote sampled data [2].

However, mixed pixels can also exist in an *An. arabiensis* model because disparate materials spectra join to form a single pixel; this phenomenon is especially prevalent in remote sensing systems with low spatial resolution. To date, spectral unmixing algorithms have proliferated in a variety of disciplines that exploit low resolution and hyperspectral data. For example, the LSE method was used by Hlavaka and Spanner [23] using AVHRR data, by Maselli on Landsat TM data [29], and by Luo *et al*., [30] on a 4-channel sensor. The SVD method was used by Herries, *et al*. [31] multispectral airborne data and by Herries *et al*. [32] on simulated Landsat and SPOT TM data. Sirkeci *et al*. [33] used restricted total least squares methods on simulated hyperspectral imagery. Another unmixing method, quadratic programming, was used by Qian [34] with AVIRIS data and by Li and Bruce [22] with handheld spectroradiometer data. Petrou and Foshi [35] used a fuzzy method with simulated SPOT HRV data, and Nascimento and Dias [36] used independent component analysis with simulated AVIRIS data. Li and Bruce used Bhattacharyya distance and the discrete Haar wavelet transform for dimensionality reduction preprocessing [22]. Keshava [37] used the spectral angle mapper (SAM) metric and generalized likelihood ratio test on HYDICE data. Zhu and Eastman [38] used minimal residual and probability guided algorithms. Ma *et al*. [39] used a genetic algorithm on HYMAP data. Riedmann and Milton [40] used a supervised maximum likelihood selection algorithm on HYMAP and CASI data. Chein-I *et al*. [41] used divergence and band decorrelation algorithms on HYDICE data. Haertel and Shimabukuro [42] estimated multiple land cover components such as vegetation, bare soil and shaded areas from Landsat TM data and then successfully derived unknown components' endmembers in the low spatial resolution TerraMODIS image by using a linear spectral mixing model. Zurita-Milla *et al*. [43] inverted the linear spectral mixing model to obtain MERIS endmembers based on the known fractional coverages of each pixel also from a Landsat TM classification as well. However, an *An. arabiensis* habitat pixel extracted from low resolution data may still contain mixtures of unidentified habitat components after spectral decomposition; therefore, the selected single pixel extracted may not present the field-sampled data accurately. For example, multitemporal Landsat TM has been used to estimate paddy rice fields [44, 45], but its spatial resolution (30m) makes it difficult to correctly identify *An. arabiensis* habitats [46]. Even Multispectral thermal Imager (MTI) data at 5m cannot provide an optimum scale to segment and classifiy an individual anopheline habitat pixel [47].

A spatially quantitative pixel assessment for targeting highly productive *An. arabiensis* habitats from gridded QuickBird visible and near infra-red (NIR) data (www.digitalglobe.com) at 0.61m spatial resolution may help direct resources for implementing larval control strategies in riceland environments. Variograms constructed using QuickBird data have been used to describe the degree of spatial dependence and stochastic processes in spatiotemporal-sampled vector mosquito habitat explanatory predictor variables [1]. Empirical semiovariograms have been efficiently computed for sub-meter resolution data on a regular grid from the spectral density function using the fast Fourier Transformation (FTT) and smoothed periodogram [5]. For remote identification of georeferenced vector mosquito habitats the first step is often to construct a discrete tessellation of the region [17]. Unfortunately, the FTT and the empirical semiovariogram cannot be applied on most geoecological spatiotemporal-sampled datasets of riceland *An. arabiensis* habitats as often these data are remotely captured by unequally spaced sampling matrices. In this research we used an Ordinary interpolator and a digitized grid-based algorithm overlaid on QuickBird visible and NIR data for predicting productive *An. arabiensis* habitats based on spectrally extracted pixel endmember estimates. Orthogonal grid-based algorithms overlaid on satellite data of riceland areas will fail to capture physical or man-made structures (e.g. paddies, canals, berms) at these habitats but, digitizing georeferenced *An. arabiensis* habitats in ArcGIS® can convert a polygon into a grid cell which may conform to riceland habitat boundaries. Our assumption was that an Ordinary kriged-based model could use data generated from the sub-meter digitized gridded QuickBird data for estimating optimal spectral endmember predictors of productive *An. arabiensis* habitats based on spatiotemporal field-sampled count data and their underlying autocorrelation structures.

Ordinary kriging is a geostatistical approach to modeling which relies on the spatial correlation structure of the data to determine the weighting values which includes rigorous modeling of correlation between data points for determining the estimated value at an unsampled point. In this research the Ordinary kriged-based model was constructed by incorporating the covariance of the random function on a weighted moving average interpolation for spectrally predicting productive An. *arabiensis* habitats by combining the prior distribution of the endmembers with a Gaussian likelihood function. The resulting posterior distribution was Gaussian with a mean and covariance that was computed from the observed habitat pixel values, their variance and the kernel matrix derived from the original distribution.

In this research, we kriged a spectral endmember reference signature extracted from a productive *An. arabiensis* habitat based on field-sampled count data sampled in Karima rice-agro village complex in the Mwea Rice Scheme, Kenya, using the Li-Strahler geometric-optical model. Topographic effects on bidirectional and hemispherical reflectance can be calculated with a geometric-optical model [20] Multiple models of vegetation canopy reflectance have been quantified using Li-Strahler geometric-optical models. For example, a simple bidirectional reflectance model suitable for homogeneous plant canopies was developed by Gao [19] to simulate directional changes of spectral reflectance from tall grass canopies which was then compared with field-sampled measurements. A modified two-stream approach was used in the model by considering the transfers of upward reflected and diffuse radiation in different viewing directions, downward direct and diffuse radiation, and their dependences upon canopy leaf orientation. An analytical solution was derived to allow a fast computation of directional reflectance based on canopy parameters or a retrieval of canopy parameters based on reflectance measurements. Modeled spectral reflectance was generally in a good agreement with measurements, and effects of inhomogeneous canopy characteristics such as hot "spot" and of soil reflection. Both modeled and observed reflectance estimates revealed strong backscattering and relatively weak forward scattering in the red band, but a more symmetric distribution in the NIR band. The directional changes of reflectance in both red and NIR bands were dominated by primary scattering; the magnitudes of reflectance in the NIR band were largely influenced by secondary scattering. Furthermore, model simulations for three theoretical leaf angle distributions, planophile, spherical, and erectophile, revealed that hemispherical reflectance (i.e., surface albedo) could be overestimated by nadir reflectance at a small solar zenith angle and underestimated at a large solar zenith angle. They also found that the error can be reduced to a minimum at an optimal solar zenith angle which had the same magnitude for the three leaf angle distributions but this uncertainty probability estimate decreased with increasing canopy leaf area index. The final model vegetation spectral index varied by 10–24 % with viewing angle for a given leaf angle distribution and canopy leaf area index.

Biophysical parameters quantified from ground reflectance data of an *An. arabiensis* habitat using the Li-Strahler model may determine an optimal endmember habitat dataset. Furthermore, by analyzing a simple model of bidirectional reflectance estimates for *An. arabiensis* riceland habitat canopy cover with azimuthally nonuniform leaf distributions may establish sampled rice leaf inclination values using a "Lambertian-viewing" cone around the nadir for a range of solar zenith angles. If a surface exhibits Lambertian reflectance, light falling on it is scattered such that the apparent brightness of the surface to an observer is the same regardless of the observer's angle of view (i.e., the surface luminance is isotropic) [5]. In this region, rice leaf inclination and rice leaf azimuth may not materially affect the bidirectional reflectance from an *An. arabiensis* habitat which may be then effectively constant thus, generating false statistical inferences in unmixing models. According to Schowengerdt [20] under constant bidirectional reflectance viewing / illumination geometry, only the product r cos Ψ may be inferred for a complete habitat canopy and r $\cos^2 \Psi$ ΔL for a sparse canopy. The r cos Ψ and ΔLcos Ψ are very basic canopy-related parameters, influencing canopy bidirectional reflectances. This would also apply to more complicated *An. arabiensis* habitat products that involve an additional term which may be a function of the solar zenith angle, rice leaf azimuth, and rice leaf zenith angle, where r is rice leaf reflectance, Ψ is zenith angle of rice leaf normal and ΔL is rice leaf area per unit horizontal area. Thus, the rice leaf reflectance r associated to a spatiotemporal-sampled *An. arabiensis* habitat cannot be inferred. The rice leaf reflectance r may only be determined if Ψ is inferred in the *An. arabiensis* habitat model which may only be possible outside the "Lambertian-viewing" cone. These limitations may be predicated on the Lambertian leaf reflectance characteristics computed from the *An. arabiensis* habitat canopy structures and may not apply when rice leaf reflection has a pronounced specular component. To assess the limitations of field measurements in a planar scan, we can then reformulate an An. *arabiensis* habitat model using bidirectional reflectance estimates by a change in the coordinate system, expressing leaf orientation by two projection ratios of the rice leaf area. We may then be able to establish viewing solely in the solar principal plane. Commonly, the projection ratio q (i.e., projection of the rice leaf area on the principal plane divided by that on the horizontal plane) cannot be inferred in an *An. arabiensis* habitat model [1]. Characterization of the *An. arabiensis* habitat "hot spot", for which the view direction is close to the direction of illumination, thus would have to be quantified using the expression for the bidirectional reflectance in terms of the width of the cone of the view angles around the direction of illumination. Determination of this cone width may, however, require narrow field-of-view measurements in the *An. arabiensis* habitat "hot spot" region, after other canopy model parameters have been determined from measurements outside the "hot spot".

In this research, bidirectional and hemispherical reflectance (i.e., surface albedo) estimates from a geometric-optical model were used to quantify the effects of

topography on both the bidirectional reflectance distribution function (BRDF) and the hemispherical reflectance of a productive *An. arabiensis* habitat. Recently, the Li-Strahler geometric-optical model has been extended to accommodate sloping terrain in its computation of the BRDF and spectral surface albedo [5]. The model treated the rice-vegetation canopy of the sampled *An. arabiensis* habitat as an assemblage of partially illuminated objects of spheroidal shape. The model also employed geometric optics and Boolean set theory for generating the proportion of sunlit or shadowed canopy and background as functions of view and illumination angles. We then used a convex geometrical algorithm to check the pixel constituent material to determine their respective spectral endmember and abundance estimates. Thereafter, an Ordinary kriged-based interpolator was employed to forecast other prolific *An. arabiensis* habitats from the pixel spectral components quantitatively decomposed by the unmixing algorithm. If a spatial algorithm processes a pixel by utilizing means or covariances then the algorithm is statistical [7]. Our assumption was that predictive spatial autoregressive model generated using a quantitative robust QuickBird endmember pixel assessment may target highly productive *An. arabiensis* habitats for implementing control strategies in riceland environments. Therefore, our research objectives were to: (1) to construct a regression model to determine productive habitats based on spatiotemporal field-sampled count data (2) quantify bidirectional and hemispherical reflectance using a Li-Strahler model for estimating the spectral fractions of individual habitat surface components present in a 0.61m image pixel; (3) validate the endmembers generated use a convex geometrical algorithm and, (4) generate optimal predictors from an Ordinary interpolator using the productive habitat pixel radiance estimates for remotely forecasting prolific *An. arabiensis* habitats in Karima rice-village complex, Mwea, Kenya.

2. MATERIAL AND METHODOLOGY

2.1. Study Site

The study was conducted 112 km northeast of Nairobi, Kenya in Karima rice-village complex within the Mwea Rice Scheme. Mwea occupies the lower altitude zone of the Kirinyaga District, in an expansive low-lying, formally wet savannah ecosystem. The scheme is situated on the foot hills of Mount Kenya at 37^{o} 20'E and 0^{o}, 41'S. The Mwea Rice Scheme is located in the west central region of Mwea Division and covers an area of approximately 13,640 hectares. More than 50% of the scheme area is used for rice cultivation. The remaining area is used for subsistence farming, grazing, and community activities. The mean annual precipitation is 950 mm, with maximum rainfall occurring in April–May and October–November. The average temperatures range from 16°C to 26.5°C. Relative humidity varies from 52% to 67%. According to the 2009 Kenyan national census, the Mwea Rice Scheme has a population of 150,000, occupying 25,000 households. The Karima study site is located at the central-west region of the scheme and has 158 homesteads, with approximately 950 residents.

In Mwea, the beginning of each cropping or growing cycle is scheduled according to the water availability through the irrigation water distribution scheme. The schedule of individual farmers rice planting also differs within this time when water is available. Most fields are cultivated once a year, although some farmers cultivate a second crop. The typical cultivation cycle includes a sowing–transplanting period (June–August), a growing period (August–November) and a post-harvest period (November–December). The second crop is cultivated prior to the short rainy period between January and May. The duration of the rice cycle varies between 120 and 150 days, depending on the rice variety planted. After harvesting, active mosquito habitats may persist in shallow puddles left after harvest [1].

The start of the rice growing cycle begins with paddy preparation by pre-flooding the paddy with about 12cm of water. Farmers sow rice seeds into a small section (nursery) of their paddy to obtain single stemmed plants for transplanting throughout the paddy. For transplanting, the number of seedling plants is 3-4 seedlings per hill with a planting density is 20-30 hills per m^2. Fertilizers in the form of Ammonium sulphate and triple super phosphate are applied at a rate of about 50 kg/ha and 125 kg/ha, respectively, about five days before transplanting. Studies by Mwangangi *et al.* [48] showed that dose amount of ammonium sulfate accounted for up to 40% mortality rate and one week delay in development time to *An. arabiensis* larvae. The actual amount of water used by farmers for land preparation and during the crop growth period, however, is much higher than the actual field requirement. Paddy farmers in Mwea often store water in their fields as a back-up safety measure due to the unreliability in supply of water for irrigation. This leads to a high amount of surface runoff, seepage and percolation, accounting for about 50–80 percent of the total water input to the field creating active mosquito habitats.

2.2. Habitat Sampling

Base maps were generated in ArcGIS® of the study site (Fig. **1**). Field sampling was performed from June 2009 to June 2010. Permanent, semi-permanent and temporary and *An. arabiensis* habitat sites in the study site were classified and mapped using a CSI-Wireless Max differentially corrected global positioning system (DGPS) receiver employing a OmniStar L-Band satellite signal with a positional accuracy of less than 1m (Clarke Mosquito Control Products, Inc. 159 N. Garden Avenue. Roselle, IL 60172) [49].

Water bodies were inspected for mosquito larvae using standard dipping techniques with a 350-ml dipper to collect the mosquito larvae [50]. The number of dips per habitat was 20. All data from the habitat characterization of each larval habitat was recorded on a field sampling form. Larvae and a sample of water from each larval habitat were placed in whirl-pack bags and transported to the Mwea Research Station for further processing. All 3^{rd} and 4^{th} instar larvae were immediately preserved in 95% ethanol and later identified morphologically to species using taxonomic keys [51, 52]. The 1^{st} and 2^{nd} instars were reared in plastic pans

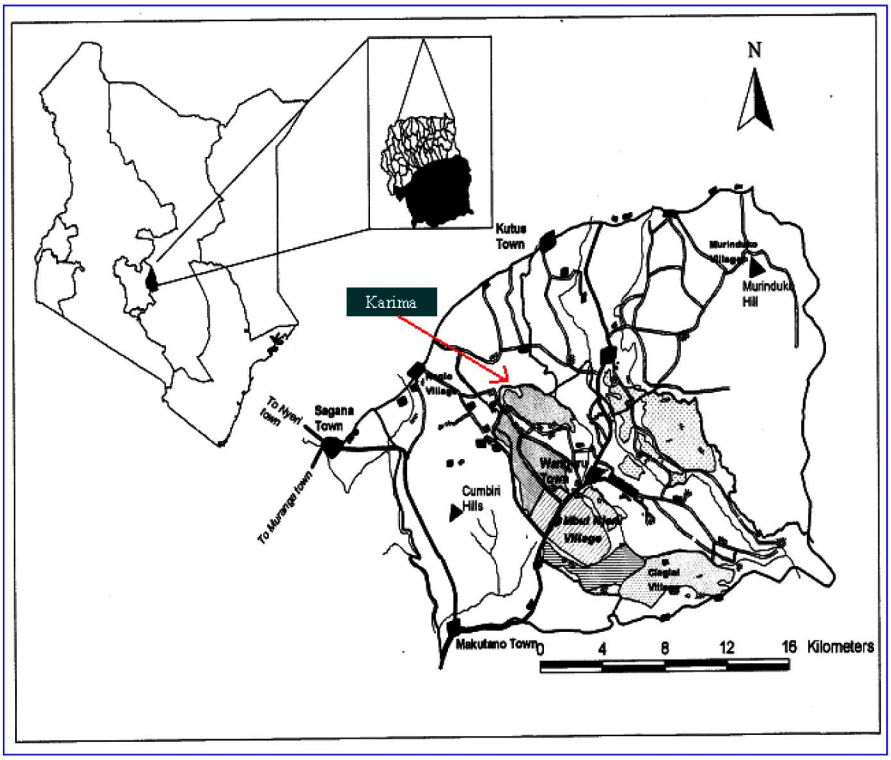

Fig. (1). Base map of the Karima study site.

under semi-field conditions and those that survived to 3rd instar were also preserved and identified morphologically to species. The pupae were kept in mosquito emergent cages (Bioequip Products Inc) and the resultant emergent mosquitoes were identified morphologically.

2.3. Aquatic Habitat Characterization

Meteorological data was acquired from a Davis Instruments 6153-IP Ethernet Wireless Weather Station placed in a secured homestead within the study site. Environmental variables recorded for each habitat were number of aquatic animals, depth, distance to the nearest house, canopy coverage, shade, and turbidity. Distance to the nearest house was measured with a tape when it was shorter than 100 m. When the distance exceeded 100 m, it was estimated visually. The distance to the nearest house was categorized into 7 classes (e.g., 1: 50–100 m, 2: 101-200 m, and so on, and 7 for distances greater than 600 m). Canopy cover was defined as the amount of terrestrial vegetation and other objects in the habitat. The number of individuals of each family identified were counted and recorded. Shade coverage of a habitat was measured in percentage of water surface covered by placing a square frame ($1m^2$) with grids ($100\ cm^2$) above the sampled habitat.

2.4 Remote Sensing Data

Raster image data from the DigitalGlobe QuickBird satellite service were acquired for the study site for the periods of: 15 July 2006, 1 September 2006, 1 February 2007, and 15 July 2007 within the study site area. The QuickBird image data were delivered as pan-sharpened composite products in infra-red (IR) colors. The clearest, cloud-free images available of the contiguous sub-areas of the study site were used to identify land cover and other spatial features associated with *An. arabiensis* habitats. The Order Polygon contained 5 vertices consisting of longitude/ latitude (decimal degrees) geographic coordinates using a WGS-84 ellipsoid. The satellite data contained 64 km^2 of the land cover in the study site.

The QuickBird imagery was classified using the Iterative Self-Organizing Data Analysis Technique (ISODATA) unsupervised routine in ERDAS *Imagine* v.8.7™ (ERDAS, Inc., Atlanta, Georgia). Unsupervised classifications are commonly used for the identification of land covers and mosquito habitats associated with intermediate hosts and disease vectors [53-55]. QuickBird collects data using an 11-bit dynamic range. This allowed 2^{11} or 2048 possible intensity values for each pixel. Because computers cannot read 11-bit data, Digital Globe delivered the data in an 8 bit format. The 8-bit data required that QuickBird 11-bit data be compressed as such data spanning 2,048 digital values was rescaled to 256 values.

2.5. Habitat Mapping

Each *An. arabiensis* habitat in the study site with their corresponding data attributes was entered into the VCMS™ relational database software product (Clarke Mosquito Control Products, Roselle, IL). The VCMS™ database supports a mobile field data acquisition component module, called Mobile VCMS™ that synchronizes field data from

industry standard Microsoft Windows Mobile™ devices and can support add-on GPS/DGPS data collection [56]. Mobile VCMS™ and its corresponding FieldBridge® middleware software component were used to support both wired and wireless synchronizing of field data from the habitat monitoring. Field data collected with Mobile VCMS™ was synchronized directly into a centralized VCMS™ relational database repository. Additional geocoding and spatial display of the ecological data was handled using the embedded VCMS™ GIS Interface Kit™ that was developed utilizing ESRI's MapObjects™ 2 technology. The VCMS™ database supported the export of all field data, using any combination of georeferenced *An. arabiensis* habitats at the study site, in order to further process and spatially display specific data attributes in a stand-alone desktop GIS software package (ESRI ArcGIS).

The VCMS™ database plotted incoming DGPS ground coordinates of the sampled *An. arabiensis* habitats collected using a Trimble Recon® X PDA (400MHz Intel PXA255 Xscale CPU) and Mobile VCMS™ software. The data was divided into four equal quadrant sections of 120 Mb per quadrant as the total raw QuickBird image size was 655 Mb. This imagery was further processed into tiles making it suitable for display on the Recon X®. To process the imagery into smaller manageable datasets, we used pixel grids to draw a minimum-bounding rectangle around the image-oriented map which was then divided into tiles. We specified an 8k x 8k pixel based on the amount of data in each tile using ERDAS *Imagine* 9.2® software.

ESRI ArcPad® software was installed on the Recon X® and we created an ArcPad folder on the C drive of our field computer (laptop PC) to connect and synchronize data with ArcPad on the PDA. The ArcPad® Data Manager for ArcGIS Desktop, the Datum Configuration Tool and the ArcPad® Deployment Manager were copied into this folder, along with other ArcPad® modules. This software was used to display and manipulate the QuickBird imagery on the Recon PDA field computer.

2.6. Grid-Based Algorithm

A polygon layer outlining each *An. arabiensis* habitat within a 1 km buffer was created by digitizing the QuickBird imagery in ArcGIS ®. Each cell within the matrix contained an attribute value, as well as habitat location coordinates, and was joined relationally to other databases. The spatial location of each cell was implicitly contained within the ordering of the matrix. The habitats were then characterized in relation to the ecological attributes sampled of an aquatic habitat. Each habitat/polygon was assigned a unique identifier. Field attribute tables were then linked to the polygons. The polygons were used to define the sampling frame, which extended to include a 1 km buffer from the external boundary of the Karima rice-village study site. A mark-release-recapture study in an area near Bamako, Mali, showed that *An. arabiensis* generally does not disperse further than 1 km [57]. In this research, digitized QuickBird grid cells were stratified based on LULC transition throughout the rice cycle and defined as: ploughed (paddy preparation), flooded, post-transplanted, tillering, flowering and fallow/post-harvest which were defined as:

1) Ploughing: This LULC was the field preparation prior to transplanting of rice seedlings.
2) flooding: This classification comprised of areas of intensive land use with much of habitat covered by water.
3) Post-transplanting: This rice-cycle time period was following rice seedlings transplanting.
4) Tillering: This class extended from the appearance of the first tiller until the maximum tiller (5-9) number was reached. During this phases stem elongation occurred and the tillers continued to increase in number and height, with increasing ground cover and canopy formation.
5) Flowering and maturation: In this LULC the stage plants stopped growing and oriented towards the development of the panicles and plants senesce and their water content dropped. The flowering/maturation phase included the panicle initiation, booting, heading and flowering stages.
6) Fallow/post-harvest: This period occurred after harvesting when the land was left bare waiting the next crop cycle (Fig. **2**).

Overall accuracy and class-specific user and producer accuracies were calculated for each of the resultant land cover classes. For each mapping region, stratified sampling formulas were applied to estimate the error matrix cell proportions [58], and consequently, the estimates of overall and class-specific user's and producer's accuracy [59]. In this research, the producer's accuracy related to the probability that a land cover class was correctly mapped and measured using the errors of omission (1 - producer's accuracy). In contrast, the user's accuracy indicated the probability that a sampled *An. arabiensis* habitat from the land cover map actually matched the information from the georeferenced datasets by measuring the error of commission (1- user's accuracy). Accuracy results were computed by weighting the proportions of each land cover, within the study site, against total land cover area used in the sampling frame. Specifically, the overall accuracy $\left(\hat{P}\right)$ and producer's accuracy $\left(\hat{P}_{Aj}\right)$ were estimated using post-stratified formulas [58]. Post-stratified estimators use the known pixel totals for each land-cover class (Ni+), treating the sample as a stratified random sample of ni+ pixels from the Ni+ pixels in that class [2]. In this research the user's accuracy $\left(\hat{P}_{Ui}\right)$ was based on the simple random sampling formulas:

$$\hat{P} = \frac{1}{N}\sum_{k=1}^{q}\frac{N_{k+}}{n_{k+}}n_{kk} \quad (1)$$

$$\hat{P}_{Ui} = n_{ii}/n_{i+} \quad (2)$$

$$\hat{P}_{Aj} = \frac{\left(N_{j+}/n_{j+}\right)n_{jj}}{\sum_{k=1}^{q}\left(N_{k+}/n_{k+}\right)n_{kj}} \quad (3)$$

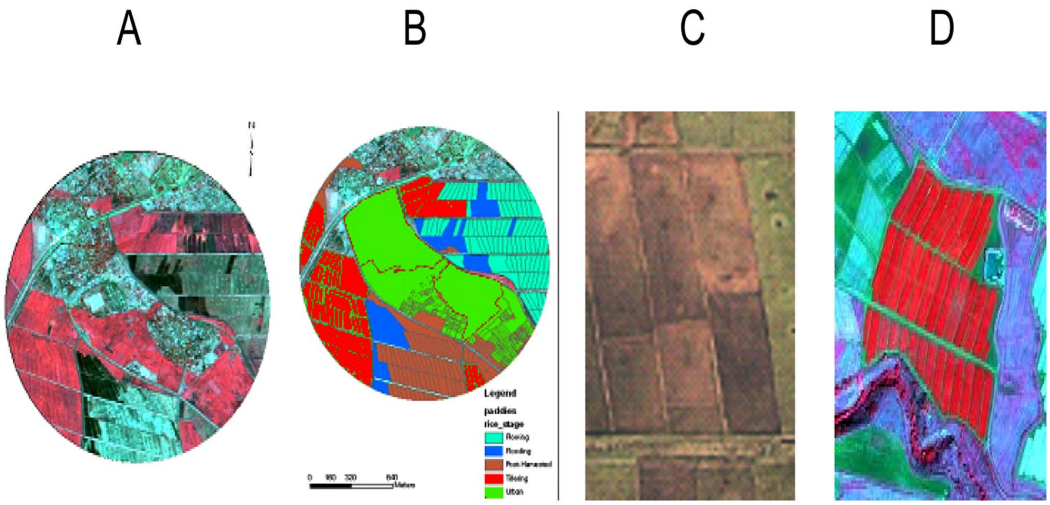

Fig. (2). Isolating paddy preparation *An. arabiensis* habitat pixels form a QuickBird land use cover map in the Karima study site.

A. QuickBird visible and NIR data of study site within a 1 km buffer.

B. *An. arabiensis* habitats classified by land use land cover (LULC) within a digitized grid-based algorithm .

C. Paddy-preparation *An. arabiensis* habitat.

D. Pixel's of paddy preparation *An. arabiensis* habitat.

The spatial distribution of the sampled habitats were created as a layer and overlaid on the land cover layer, and the number of *An. arabiensis* habitats in each land cover class was calculated using Zonal stat, an avenue script in ArcView® [Earth Systems Research Institute (ESRI), Redlands, California]. The chi-square analysis for the sampled field and remote explanatory predictor variables was used to examine whether there were significant differences in proportions of positive and negative sampled aquatic habitats in each grid cell located in different land cover classes.

We tested the kappa statistics using different significance levels and the LULC covariates. In general, (kappa) independent significant tests the (alpha) level of null hypotheses the probability that there is no significant differences is $(1\text{-}(\text{alpha})^{(\text{kappa})})$ (1). In our analyses, the important factors for generating kappa probability statistics between the LULC covariates was to discover if the codes generated were equiprobable and to determine if the marginal probabilities for the classification data were similar or different. In this research we used equally arbitrary guidelines to characterize kappas over .75 as excellent, .40 to .75 as fair to good, and below .40 as poor. Three matrices were involved, the matrix of observed scores, the matrix of expected scores based on chance agreement and the weight matrix. Weight matrix cells located on the diagonal (upper-left to bottom-right) represented agreement and thus contained zeros. Off-diagonal cells contained weights indicating the seriousness of that disagreement. The equation for weighted κ was: $k = 1 - \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}x_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}m_{ij}}$ where k=number of codes and w_{ij}, x_{ij}, and m_{ij} were elements in the weight generated from the LULC classification.

2.7. Object Oriented Classification

ENVI® (www.ittvis.com). spectral tools were used to analyze the QuickBird data, in order to confirm the location of varying states of rice crops. Jacob *et al.* [1] used ENVI for supporting Input File QuickBird functions including GeoTIFF, NITF and QuickBird Tile Product (.til) which segregated homogenous pixels of anopheline habitats. A prerequisite to classification is image segmentation, which is the subdivision of an image into separated regions [5]. Image objects resulting from segmentation represent image object primitives, serving as information carriers and building blocks for further invasive image classification or other segmentation processes [2]. In ENVI 4.5®, a spectrum plot, known as a *Z*-profile of the pixel under the cursor was run through all bands of the QuickBird image of the Karima study site. The basic workflow involved importing the data collected in the field into a spectral library. The spectral library then used the Endmember Collection workflow to perform a supervised classification based on the ecologically sampled *An. arabiensis* habitat data. Binary Encoding, Spectral Angle Mapper (SAM) and Spectral Feature Fitting were used to rank and match any unknown pixel spectrum to the *An. arabiensis* pixel materials in the spectral library. The Spectral Angle Mapper Classification (SAM) is an automated method in object-oriented classifications for directly comparing image spectra to a known spectra (usually determined in a lab or in the field with a spectrometer) or an endmember [20].

In this research SAM was used for comparing image spectra to the spectral library. The algorithm determined the similarity between spectra by calculating spectral angle and then treated them as vectors in a space with dimensionality equal to the number of QuickBird bands. This method however was insensitive to the *An. arabiensis* habitat

illumination estimates since the algorithm used only the vector direction and not the vector length. A smaller angle means a closer match between the two spectra and the pixel is identified as the field spectrum [2]. Because it used only the "direction" of the spectra, and not their "length," the method was insensitive to the unknown gain factor, and all possible illuminations were treated equally. SAM determined the similarity of any unknown spectrum *t* to a reference spectrum r, by applying $\alpha = \cos^{-1}\left[\frac{\vec{t} * \vec{r}}{\|\vec{t}\| * \|\vec{r}\|}\right]$ which was written as $\alpha = \cos^{-1}\left[\frac{\sum_{i=1}^{nb} t_i r_i}{\left(\sum_{i=1}^{nb} t_i^2\right)^{1/2}\left(\sum_{i=1}^{nb} r_i^2\right)^{1/2}}\right]$. For each reference spectrum chosen in the library, a spectral angle was determined. This angle in radians was assigned to the corresponding pixel in the output SAM image. The angles were subjected to thresholds to determine their class.

The image endmembers of the *An. arabiensis* habitats were then extracted from ENVI's spectral library. Several spectra corresponding to the same green rice-vegetation type over different backgrounds in the *An. arabiensis* habitats had to be included, since multiple scatterings between tillers and a bright soil background increased the QuickBird NIR reflectance of the tillers. After the calibration coefficients had been determined the image was converted to match the library. Analogously, the reference endmembers spectra in the library were transformed into the endmembers spectra of the image.

We employed a method to empirically assess the classification accuracy of the QuickBird visible and NIR data by selecting LULC classes and comparing them with the reference data. In this research we used a random selection of pixels in each LULC class to assess the satellite classification accuracy. QuickBird class representative pixels were selected and compared to a reference training dataset. The delineation of training sites representative of land cover types for identifying anopheline habitat is most effective using the spectral properties of LULC classes [17]. During the segmentation procedure, image objects were generated based on several adjustable criteria of homogeneity such as color, shape, and texture. Data pre-processing involved converting DN to radiance, atmospheric correction using FLAASH^{TM}, and co-registration. Image classification was done using the object-oriented approach. FLAASH^{TM} which generated a spatial model (.gmd file) that converted the image's DN to at-sensor radiance and computed at-sensor reflectance while normalizing solar elevation angle. The equation was as follows:

$$\rho_{BandN} = \frac{\pi (L_{BandN} * Gain_{BandN} + Bias_{BandN}) * D^2}{E_{BandN} * (COS((90-\theta) * \pi / 180))}$$

where,

ρ_{BandN} = Reflectance for Band N

L_{bandN} = Digital Number for Band N

D = Normalized Earth-Sun Distance

E_{bandN} = Solar Irradiance for Band N

The reference data in this research was the "ground truth" data of the sampled habitat predictor variables. Selected random pixels from the thematic map were then compared to the reference data. The estimated accuracy for LULC data was calculated using:

$$P\left\{-z_{\frac{\alpha}{2}} < \frac{x - n\theta}{\sqrt{n\theta(1-\theta)}} < z_{\frac{\alpha}{2}}\right\} = 1 - \alpha$$

where, *x* was number of correct identified pixels, *n* was total number of pixels in the sample, θ was the map accuracy, and (1-α) was a confidence limit. The QuickBird classifier identified *An. arabiensis* habitats such as water bodies larger than 0.4 ha with generally satisfactory results (92.1%) with a lower detection limit of 0.1 ha.

We expected the *An. arabiensis* habitat larval/pupal count in the study site to follow a Poisson distribution, as was the case in previous research in Kenyan irrigated riceland areas [1, 61, 62]. Therefore, we used the count data and standard deviations of the log-number of larval/pupal counts collected in the study site to determine sample size requirements. We applied a sampling intensity formula for determining the number of samples to collect from an infinite population $n = (ts/E)^{\wedge 2}$,where t = *t* value ($t \approx 2$), s = the standard deviation of log-larval/pupal count values observed in the study site (s= 0.889) and E was the desired half-width of the confidence interval around the mean expressed in same units as standard deviation (E= ln(1.25) [63]. Applying this formula we determined 152 samples were required. We overlaid vector images of the sampling scheme with the raster image of the study site to identify areas of interest (AOI) within the sampling frame for field crews to visit. All potential aquatic larval habitat sites were identified and data relative to species composition and abundance, predators, water quality and environmental parameters were collected longitudinally by field crews.

2.8. Regression Analyses

Field data parameters were entered in Microsoft Excel files and analyzed using and SAS 9.2® (SAS inc. Carey, NC, USA). Before analyses the data was tested for collinearity using design matrix from a Poisson regression model and run through SAS PROCREG/ Variant Inflation Factor (VIF) procedure. Variance inflation factors are a measure of the multicolinearity in a regression design matrix (i.e., the independent variables). Multicollinear *An. arabiensis* explanatory variables are difficult to analyze because their effects on the response can be due to either true synergistic relationship among the explanatory predictor variables or a spurious correlation [1]. To avoid over parametrization of the model we added the VIF for each spatiotemporal-sampled *An. arabiensis* variable.

We considered the following linear model with *k* independent variables: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$. The standard error of the estimate of β_j is $\sigma(X'X)^{-1/2}_{j+1,\, j+1}$, where *X* was the regression design matrix — a matrix such

that $X_{i, j+1}$ was the value of the j^{th} covariate for the i^{th} for the spatiotemporal-sampled *An. arabiensis* observations, and $X_{i, 1}$ equaled 1 for all i.

In this research the variance was expressed as : $\text{var}\left(\hat{\beta}_j\right) = \frac{\sigma^2}{(n-1)\overline{\text{var}}\left(X_j\right)} \cdot \frac{1}{1-R_j^2}$, where R_j^2 was the multiple R^2 for the regression of X_j of the sampled *An. arabiensis* habitat covariates. This identity separated the influences of several distinct factors on the variance of the coefficient estimate. The remaining term, $1/(1-R_j^2)$ was the VIF which reflected all factors that influenced the uncertainty in the coefficient estimates. The VIF equals 1 when the design matrix is orthogonal and is greater than or equal to 1 when the design matrix is not orthogonal [64]. The VIF was invariant to the scaling of the *An. arabiensis* predictor variables, that is, we could scale each variable X_j by a constant c_j without changing the VIF. The VIF was calculated in three steps. We calculated k different VIFs, one for each X_i by first running an ordinary least square regression that used X_i as a function of all the other sampled predictor variables in the first equation. We used the equation $X_1 = \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_k X_k + c_0 + e$, where c_0 was a constant and e was the error term. We also used VIF= 1/1-R_i^2. We then calculated the VIF factor for $\hat{\alpha}_i$ with the following formula: $VIF = \frac{1}{1-R_i^2}$ where $R^2{}_i$ was the coefficient of determination of the regression equation. We then analyzed the magnitude of collinearity by considering the size of the $VIF\left(\hat{\alpha}_i\right)$. The var($X_j$) revealed greater variability in the spatiotemporal-sampled *An. arabiensis* explanatory covariates which lead to proportionately less variance in the coefficient estimates.

ANOVA test was also used to compare the differences in the *An. arabiensis* larval/pupal abundance among the rice stages. The regression line concept we used was $\left(y_i - \bar{y}\right) = \left(\hat{y}_i - \bar{y}\right) + \left(y_i - \hat{y}_i\right)$ $(\boldsymbol{y_i - \bar{y}}) = (\boldsymbol{\hat{y}_i - \bar{y}}) + (\boldsymbol{y_i - \hat{y}_i})$, where the first term was the total variation in the response y (total immature count of *An. arabiensis*), the second term was the variation in mean response based on the sampled parameters and the third term was the residual value in the model estimates. Squaring each of these terms and adding over all of the sampled observations rendered the equation $\sum\left(y_i - \bar{y}\right)^2 = \sum\left(\hat{y}_i - \bar{y}\right)^2 + \sum\left(y_i - \hat{y}_i\right)^2$ $\sum(\boldsymbol{y_i - \bar{y}})^2 = \sum(\boldsymbol{\hat{y}_i - \bar{y}})^2 + \sum(\boldsymbol{y_i - \hat{y}_i})^2$. This equation was then written as SST = SSM + SSE, where SS was notation for sum of squares and T, M, and E were the notation for the model, and the residual variance error, respectively. In this research, the square of the sample correlation was equal to the ratio of the estimates while the sum of squares was related to the total sum of squares: r^2 = SSM/SST. This formalized the interpretation of r^2 as explaining the fraction of variability in the sampled *An. arabiensis* habitat parameters explained by the regression model. The sample variance $s_y{}^2$ was equal to $\sum\frac{\left(y_i - \bar{y}\right)^2}{n-1}$ $\sum\frac{(\boldsymbol{y_i - \bar{y}})^2}{n-1}$, which in turn was equal to the SST/DF (degrees of freedom), the total sum of squares divided by the total DF. A linear regression equation was constructed using the mean square model (MSM)= $\sum\frac{\left(\hat{y}_i - \bar{y}\right)^2}{l}$ $\sum\frac{(\boldsymbol{\hat{y}_i - \bar{y}})^2}{l}$, which was equal to the SSM/DF. The corresponding mean square error (MSE) was $\sum\frac{\left(y_i - \hat{y}_i\right)^2}{n-2}$ $\sum\frac{(\boldsymbol{y_i - \hat{y}_i})^2}{n-2}$, which was also equal to SSE/DF and also the estimate of the variance about the regression line (i.e., σ^2). The MSE is an estimate of σ^2 for determining whether or not the null hypothesis is true [64]. ANOVA calculations were then displayed in an analysis of variance table. The null hypothesis for the ANOVA analyses was based on the average value of the dependent variable (i.e., larval/pupal count of *An. arabiensis*) which was the same for all sample groups tested.

Generally, ANOVA used in vector mosquito data analyses tests the null hypothesis that all the sampled parameters of the population means are equal (i.e., H_0: $\mu_1 = \mu_2 = ... = \mu_a$), by comparing two estimates of variance [1, 62]. If the null hypothesis is false, then Mean Square Between (MSB) estimates generated from a linear predictive vector mosquito habitat model is something larger than σ^2 [1] The MSE is an estimate of variance for determining whether or not the null hypothesis is true, while the second estimate MSB is based on the variance of the sample means [64]. The logic by which our computed analysis of variance tested the null hypothesis was as follows: if the null hypothesis was true, then MSE and MSB was approximately the same since they were both estimates of the same quantity (i.e., σ^2)]; however, if the null hypothesis was false then MSB was expected to be larger than MSE, since MSB was estimating a quantity larger than σ.

The ANOVA model encompassed all possible sources of variation in the spatiotemporal-sampled habitat data which allowed us to further test our research hypotheses. In this research, the significance test of the *An. arabiensis* habitat data also involved the ratio of MSB to MSE: F = MSB/MSE. The F-statistic was used to calculate the p-value. The P value or calculated probability is the estimated probability of rejecting the null hypothesis (H_0) of a study question when that hypothesis is true [64]. Traditionally for anopheline mosquito models, if $p < .05$ the null hypothesis is rejected [47]. In this research the F ratio was approximately one, which indicated MSB and MSE were about the same**.** The residual regression output also indicated that the paddy preparation stage was associated with the highest immature *An. arabiensis* count throughout the rice cycle.

We then checked for error in our regression model. If *An. arabiensis* habitat data are heteroscedastic, nonlinearly associated, or have outliers, the regression line is not a good summary of the data, and it is an error to use regression to

summarize the data [1]. To determine if measurement error in the response variable was normally distributed and had constant variance with predictors free of measurement error, we used a residual regression plot. A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis [5]. Heteroscedasticity, nonlinearity and outliers are easily visualized in a residual plot especially when using remotely-sampled explanatory covariates [20]. A residual in a predictive vector habitat distribution model is the vertical difference between the Y value of an individual sampled habitat and the regression line at the value of X corresponding to that individual sampled habitat, for regressing Y on X. That is, if there are n pairs of measurements of X and Y: (x_1, y_1), (x_2, y_2), … , (x_n, y_n), then the equation of the regression line is $y = a \times x + b$ and the vertical residual e_1 for the first datum is $e_1 = y_1 - (a \times x_1 + b)$ The vertical residual for the second datum will be $e_2 = y_2 - (a \times x_2 + b)$, and so on. The i^{th} vertical residual is then the amount by which the regression line at the i^{th} value of X misses the i^{th} value of Y—the error in using the regression line to estimate the i^{th} datum [47]. For regressing X on Y, a residual is the horizontal difference between the X value of the individual and the regression line at the value of Y corresponding to that individual [20].

Since plotting the residuals as a function of the "independent" sampled *An. arabiensis* explanatory predictor variable could determine whether the regression was computed correctly, in this research we generated a plot of the residuals against the corresponding values of the independent variable (i.e., residual plot) . It was essentially a scatterplot of the n sampled *An. arabiensis* habitat data (x_1, e_1), (x_2, e_2), … , (x_n, e_n). A residual plot for a predictive *An. arabiensis* habitat model is like a scatterplot of the original sampled data, but with $(a \times x_i + b)$ subtracted from the value of y_i for each point (x_i, y_i), $i = 1, 2, \dots, n$[1]. In this research, subtracting the regression line from the spatiotemporal-sampled habitat data removed from Y any overall average and any trend with X.

We then generated an equation of the regression line from the spatiotemporal-sampled *An. arabiensis* habitat parameters which was $y = a \times x + b$, where $a = r_{XY} \times SD_Y/SD_X$ and b = mean(Y) - a × mean(X). The vertical residuals were $e_i = y_i - (a \times x_i + b)$, or $i = 1, 2, \dots, n$. The mean of the residuals was $1/n) \times (e_1 + e_2 + \dots + e_n) = (1/n) \times (y_1 - (a \times x_1 + b) + y_2 - (a \times x_2 + b) + \dots + y_n - (a \times x_n + b)) = (1/n) \times (y_1 + y_2 + \dots y_n - a \times (x_1 + x_2 + \dots + x_n) - n \times b)$ $SD_e = (1 - r_{XY}^2)^{1/2} \times SD_Y$.) The sum in the first set of square brackets was $r_{XY} \times SD_X \times SD_Y$, and the sum in the second set of square brackets is $(SD_X)^2$, so $r_{XE} = c \times r_{XY} \times SD_X \times SD_Y - c \times a \times (SD_X)^2 = c \times (SD_X)^2 \times [r_{XY} \times SD_Y/SD_X - a]$. In this research the regression line was $a = r_{XY} \times SD_Y/SD_X$, so the term in square brackets was zero.

The regression line did not pass through all the *An. arabiensis* data points on the scatterplot exactly unless the correlation coefficient was ±1. In general, the data were scattered around the regression line. Each datum had a vertical residual from the regression line; the sizes of the vertical residuals varied. The root mean square (rms) of the vertical residuals measures in a predictive vector habitat model is the vertical distance of a datum from the regression line [1]. Thus, the rms of the vertical residuals in this research was a measure of the typical vertical distance from the spatiotemporal-sampled *An. arabiensis* habitat data to the regression line, which also was the error in estimating the value of Y by the height of the regression line. We then let the *An. arabiensis* habitat predictor variables be grouped as (x_i, y_i), i=1, … , n. The vertical residual of the ith point was y_i- (predicted y_i)$=y_i-(r \times SD_Y/SD_X \times x_i$ + mean(Y)-$r \times SD_Y/SD_X \times$ mean(X)) which was equal to $(y_i$- mean(Y)) - $r \times SD_Y/SD_X \times (x_i$ - mean(X)). The square of the ith vertical residual was $[y_i$- (predicted y_i) $]^2=[(y_i$- mean(Y))-$r \times SD_Y/SD_X \times (x_i$-mean(X))$]^2=(y_i$-mean(Y))^2-$2 \times r \times SD_Y/SD_X \times$ (x_i-mean(X))$\times (y_i$ -mean(Y)) + $(r \times SD_Y/SD_X \times (x_i$ - mean(X))^2. We then computed the sum of those squares for the *An. arabiensis* habitat data (i = 1, 2, … , n). Note that in this research $SD_Y=\left[\left[(y_1\right.\right.$-mean(Y))^2 + ($y_2$- mean(Y))^2 + … + ($y_n$- mean(Y))^2 $\left.\left.\right]/n\right]^{1/2}$, so ($y_1$- mean(Y))^2 + ($y_2$-mean(Y))^2 + ($y_n$- mean(Y))$^2=n \times (SD_Y)^2$. Also note that $r=\left[\right.$ [(x_1-mean(X))/SD_X]×[(y_1- mean(Y))/SD_Y] + [(x_2–mean-(X))/SD_X]×[(y_2-mean(Y))/SD_Y] + [(x_n-mean-(X))/SD_X]×[(y_n-mean(Y))/SD_Y]$\left.\right]/n$, so (x_1-mean(X))×(y_1-mean(Y)) + (x_2-mean(X))×(y_2- mean(Y)) + (x_n-mean(X))×(y_n - mean(Y)) = $r \times n \times SD_X \times SD_Y$. We then considered the first term in the square of the vertical residuals. We also added the corresponding terms for i=1, … , n, which had (y_1-mean(Y))^2 + (y_2- mean(Y))^2 + (y_n- mean(Y))^2 = $n \times (SD_Y)^2$. Similarly, the sum of the second terms for i = 1, …, n, generated $-2 \times r \times SD_Y/SD_X \times \left[(x_1\right.$-mean(X))×($y_1$- mean(Y)) + ($x_2$-mean(X))×($y_2$- mean(Y)) + ($x_n$-mean(X))×($y_n$- mean(Y))$\left.\right]$ = $-2 \times (SD_Y)^2 \times r \times n \times r = -2 \times n \times r^2 \times (SD_Y)^2$. The sum of the third terms for i = 1, …, n, gave $(r \times SD_Y/SD_X)^2 \times \left[(x_1\right.$-mean(X))$)^2$ + (x_2-mean(X))^2 + (x_n-mean(X))$^2\left.\right]$ = $(r \times SD_Y/-SD_X)^2 \times n \times (SD_X)^2 = n \times r^2 \times (SD_Y)^2$. The sum of the squares of the vertical residuals in the *An. arabiensis* habitat model was thus $n \times (SD_Y)^2 - 2 \times n \times r^2 \times (SD_Y)^2 + n \times r^2 \times (SD_Y)^2$ which was equivalent to the total number of sampled *An. arabiensis* habitat parameters *(n)*$\times (SD_Y)^2 \times (1-2 \times r^2+r^2) = n \times (SD_Y)^2 \times (1 - r^2)$. To get from the sum of the squares of the vertical residuals to the rms of the vertical residuals, we divided by n and took the square-root: rms (i.e., vertical residuals)$=\left[n \times (SD_Y)^2 \times (1-r^2)/n\right]^{1/2} = SD_Y \times (1 - r^2)^{1/2}$. The model revealed that the rms of the vertical residuals from the regression (the rms error of regression*)* was $(1 - r^2)^{1/2} \times SD_Y$. The rms error of regression is usually between 0 and SD_Y in a predictive vector habitat model [1]. The error estimate was zero when r was ±1 and SD_Y when r= 0. When r was ±1, and the rms of the vertical residuals was zero. The regression, thus, accounted for all of the variability of Y. When r was equal to 0, however, the regression line did not explain any of the variability of Y. In the model the regression line was a horizontal line at height mean (Y), so the rms of the vertical residuals from the regression line was the rms of the deviations of the sampled *An. arabiensis* habitat values of Y from the mean of Y, which was by definition, the SD of Y

[20]. When *r* was not zero, the regression accounted for some of the variability of Y, so the scatter around the regression line was less than the overall scatter in Y.

2.9. Inverted Geometric-Optical Model

In this research we used the Li-Strahler geometric-optical model [65] based on the assumption that the Bidirectional Reflectance Distribution Function (BRDF) was a purely geometric phenomenon resulting from a paddy preparation *An. arabiensis* habitat scene of discrete three dimensional objects being illuminated and viewed from different positions in the hemisphere. The inverted Li-Strahler geometric-optical model was used to retrieve specific spectral habitat explanatory predictor variables. The reflectance associated with a sampled habitat was treated as an area-weighted sum of four fixed reflectance components: sunlit canopy, sunlit background, shaded canopy, and shaded background. In most geometric-optical models these four components could be simplified to three: sunlit canopy–C, sunlit background–G and shadow–T [18]. In this research, the endmember spectral components were derived using G, C, T components' classes which were initially estimated by the QuickBird image using ENVI®. For inverting the model, parts of the three components represented by (Kg) was calculated using:

$$K_g = e^{-\pi \cdot M \cdot \left[\sec(\theta_i)+\sec(\theta_v)-O(\theta_i,\theta_v,\varphi)\right]} \tag{2.1}$$

$$O(\theta_i,\theta_v,\varphi) = 1/\pi\left(\sec\theta_i + \sec\theta_v\right)\left(t - \sin t\cos t\right) \tag{2.2}$$

$$\cos t = \frac{h\left|\tan\theta_i - \tan\theta_v\cos\varphi\right|}{r\left(\sec\theta_i + \sec\theta_v\right)} \tag{2.3}$$

$$M = \frac{-\ln\left(K_g\right)}{\left(\sec\theta_i + \sec\theta_v\right)\left(\pi - t + \cos t\sin t\right)} \tag{2.4}$$

$$CC = 1 - e^{-\pi \cdot M} \tag{2.5}$$

where, *I q u q* were the zenith angles of illumination and viewing, O was the average of the overlap function between illumination and viewing shadows of individual sampled habitats as projected onto the background and *j* was the difference in azimuth angle between illumination and viewing.

In our BRDF analyses, the *An. arabiensis* habitat pixel was modeled as the limit of its directional reflectance factor using $R(i,v)$: $R(i,v) = \frac{\int\int_A R(s)\langle i,s\rangle\langle v,s\rangle I_i(s) I_v(s) ds}{A\cos\theta_i\theta_v}$ where ds was a small Lambertian surface element over area A of the QuickBird pixel; $R(s)$ was the reflectance of ds; i, v, and s represented the directions of illumination and viewing based on the reflectance components, respectively. In our model $\langle .,. \rangle$ was the cosine of the phase angle between two directions; θ was the zenith angle of a direction; $I_i(s)$ and $I_v(s)$ were indicator functions, equal to one when ds was illuminated (I_i) or viewed (I_v) or zero otherwise. Solving our double integral equation revealed s that ds was integrated over the decomposed QuickBird pixel (i.e., the footprint of the sensor' s iFOV) .

In this research, there were two kinds of prominent habitat surfaces in the pixel spectra; A-background surface was represented by Lambertian reflectance G and C, respectively. As such then re-wrote equation [1] as $R(i,v) = K_g G + \frac{C}{A}\int\int_{A_c} \frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v} ds$, where $K_g = A_g/A$ which was the proportion of background spectral data illuminated and viewed generated by the imaged *An. arabiensis* habitat attributes. In this equation the union of A_g and A_c were the intersection of the dataset of *An. arabiensis* habitat surface elements which were illuminated and viewed, only when v and i coincided. The directional reflectance of the habitat scene depended also on the habitat feature reflectance related to G and C.

In our analyses we focused on the two terms of $R(i,v) = K_g G + \frac{C}{A}\int\int_{A_c} \frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v} ds$. The first term described how the sunlit background proportion proceeded to a maximum point as viewing and illumination positions in the hemisphere coincided. The second term described how the sunlit *An. arabiensis* habitat surface, composed of the Lambertian facets, became maximally exposed to view at the hotspot, while those facets on tops became dominant at large viewing zenith angles. The hot spot correlation effect refers to the observed brightening which can occur when viewing a scene from the same direction as the solar illumination [20] which for predictive vector insect habitat modeling is commonly, noted in the visible and NIR spectral regions [60-62].

We then analyzed how the first term $K_g G$ varied with illumination and viewing geometry. As in, Schowengerdt [20] we assumed that the spatial object of interest (i.e., *An. arabiensis* habitat) and its associated explanatory spectral covariates had the shape of a spheroid, with vertical half-axis equal to b, horizontal radius equal to R, and a height to the center of the spheroid h. To accommodate the spheroidal shape in the derivations of the shadowed habitat areas, we used the transformation $\theta' = \tan^{-1}\left(\frac{b}{R}\tan\theta\right)$. We solved this equation by replacing θ with the angle that generated the same shadow area for a sphere. For simplicity, we assumed that the centers of the spheroids were randomly distributed in depth from h_1 to h_2 over A. We then assumed that G and C were constant as average signatures over A_g and A_c for properly modeling K_g and $K_c = A_c/A$.

Next the equation $R(i,v) = K_g G + \frac{C}{A}\int\int_{A_c} \frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v} ds$ was used where K_g was expressed in a Boolean model and $K_g = e^{-\lambda\pi R^2\left[\sec\theta_i' + \sec\theta_v' - \bar{O}(\theta_i,\theta_v,\phi)\right]}$ where $\bar{O}(\theta_i,\theta_v,\phi,)$ was the

average of the overlap function $O(\theta_i,\theta_v,\phi,h)$ between illumination and viewing shadows of the *An. arabiensis* habitat and its associated features. Furthermore, ϕ was the difference in azimuth angle between viewing and illumination positions of the QuickBird imaged objects associated to the *An. arabiensis* habitat. To simplify the equation, we approximated the overlap function by the overlap area and center positions of the ellipses. This approximation is justified when solar zenith and viewing zenith angles are not too large [20]. In the case of long ellipsoidal shadows, however, this approximation could have overestimated the width of the *An. arabiensis* habitat hotspot in the azimuthal direction and underestimated the width of the hotspot in the azimuthal direction. To improve the accuracy and preserve the proper *An. arabiensis* hotspot width information, we developed another approximation as follows. We used the equations $\phi = 0$ *or* $\phi = \pi$. Then we considered the overlap function in the principal plane. We used W $\phi = 0$ or π as the elliptical illumination and then determined if viewing shadows were aligned in the same direction. The overlap area was approximated by an ellipse with one axis equal to the overlap length and the other equal to the *An. arabiensis* habitat width encompassing the pixel spectral components which after quantification yielded

$$O(\theta_i,\theta_v,\phi) = \frac{1}{2}\left[\sec\theta_i' + \sec\theta_v' - \frac{h}{b}\left|\tan\theta_i' - \tan\theta_v'\cos\phi\right|\right].$$

When the overlap area was zero, the *An. arabiensis* habitat hotspot effect disappeared. We found that the shape of the hotspot function based on the viewing and illumination positions diverged due to the shape and height of the spheroids. The equation $K_g = e^{-\lambda\pi R^2\left[\sec\theta_i'+\sec\theta_v'-\bar{O}(\theta_i,\theta_v,\phi)\right]}$ was helpful to understand how the shape of the *An. arabiensis* habitat governed the shape of overlap functions. Since in this research it was important to have an exact solution for overlap function on the principal plane, we used: $O(\theta_i,\theta_v,\phi) = (t - \sin t\cos t)(\sec\theta_i' + \sec\theta_v')/\pi$ where $\cos t = \frac{h\left|\tan\theta_i' - \tan\theta_v'\cos\phi\right|}{b(\sec\theta_i' + \sec\theta_v')}$. Our strategy was to generate an exact overlap function where $\theta_v' = \theta_i'$ and ϕ varied from 0 to 2π on the principal cone using a hybrid of geometric optical and radiative transfer model for remotely capturing bidirectional reflectance over the discontinuous riceland *An. arabiensis* habitat canopy.

In our model the viewing zenith was still θ_v but, the viewing direction had a different azimuth than the illumination position. Rather than computing the overlap of ellipses emitted from the *An. arabiensis* habitat, and its components, at arbitrary inclinations and distances directly, we instead fit a linear function to the diminution of the overlaps generated from the azimuth angles. We approximated $\Phi = \frac{4R}{h(\tan\theta_v + \tan\theta_i)}$ as the azimuthal cutoff of the *An. arabiensis* habitat hotspot and linearly interpolated for ϕ between 0 and Φ or π. For the case $\Phi < \pi$, we assigned $O(\theta_i,\theta_v,\phi) = O(\theta_i,\theta_v,\phi = \pi)$ for all ϕ between Φ and π. Though this approximation we were able to quantify any errors in the overlap area. We then used the residual output from the equation $O(\theta_i,\theta_v,\phi) = \frac{1}{2}\left[\sec\theta_i' + \sec\theta_v' - \frac{h}{b}\left|\tan\theta_i' - \tan\theta_v'\cos\phi\right|\right]$ to conclude that the azimuthal width of the *An. arabiensis* habitat hotspot effect was basically determined by R/h ratio, that the outward width of hotspot on the principal plane was determined by b/h ratio and that the inward width was determined by both.

We then wanted to determine the contribution of sunlit canopy habitat surface habitat explanatory predictor variables and their shading effects. In this research, the effect of sunlit canopy on the bidirectional reflectance was quantified using the second term in equation $R(i,v) = K_g G + \frac{C}{A}\int\int_{A_c}\frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v}ds$. This variation was dependent on the both the density and angular distribution of ds in the equation. Zhang *et al* [19] assumed that each object in a scene could be modeled as a sphere without mutual illumination shading between ds elements. Then the second term was approximated as: $K_c C = \frac{1}{2}\left(1 + \langle i,v\rangle\right)\left(1 - e^{-\lambda\pi R'^2\sec\theta_v}\right)C$. In this expression, the first term was the illuminated proportion of the area of a single sphere viewed at position v and illuminated at position i. This was weighted by the second term which was the proportion of the area of spheres visible from zenith angle θ_v. Since both terms varied smoothly between zero and one, this contribution to the *An. arabiensis* habitat hotspot was quite flat. For generating a spheroid, we then replaced $\langle i,v\rangle$ by $\langle i',v'\rangle$, where $\langle i',v'\rangle = \cos\theta_i'\cos\theta_v' + \sin\theta_i'\sin\theta_v'\cos\phi$.

The first term in equation was $K_c C = \frac{1}{2}\left(1 + \langle i,v\rangle\right)\left(1 - e^{-\lambda\pi R'^2\sec\theta_v}\right)C$; however, this equation ignored the problem of mutual shading of the *An. arabiensis* habitat canopy components. This problem was handled by multiple integration, in which the mutual shadowing of canopies and other associated objects were treated in the same way as the mutual shading of leaves, for example. Our objective in this research, however, was to derive a simple approximation to describe the effect of the habitat canopy cover covariates based on collections of individual discrete reflectance surface values. To carry this out, we developed an approach that applied one-stage geometric optics to deal with the relationship between the sub-pixel endmember reflectance spectra of the *An. arabiensis* habitat surface components that were mutually shaded in the illumination direction and the parts mutually shaded in the view direction.

2.10. Linear Spectral Unmixing Based Technique

We then used a linear spectral unmixing model to calculate the percentages of the individual *An. arabiensis* habitat pixel components contained in the QuickBird image.

The model assumed that the reflectances (S) of each pixel were a linear combination of endmembers (R), which were the reflectance spectra for each habitat component. The general equations used in this research were:

$$S_j = \sum_{i=1}^{m} K_i R_{i,j} + v_j \qquad j = 1, 2, ..., p \qquad (2.6)$$

$$1 = \sum_{i=1}^{m} K_i \qquad K_i >= 0 \qquad (2.7)$$

where m was the number of components, m was the three components of C, G, and T; p is the number of image bands; K was the fractional abundance of each paddy preparation *An. arabiensis* habitat component within the pixel and v was the residual for each QuickBird visible and NIR band.

The fractions of the *An. arabiensis* habitat endmember component within the QuickBird pixel were revealed. In practice, for deriving the endmembers, we need a test image with at least *n* pixels, and the number of *n* must be more than the number of components [2]. Thus, in this research, we used the equation $S_j = \sum_{i=1}^{m} K_i R_{i,j} + v_j$ e which was more conveniently expressed in matrix notation as $V_j = S_j - K * R_j$ where;

V_j : *n*-dimensional vector of the residuals in band j;

S_j : *n*-dimensional vector of the pixels' reflectance in band j;

K : *n* x *m* matrix of the fractions;

R_j : *m*-dimensional vector of the components' reflectance in band j.

We saught a set of numerical values for the unknowns in R_j such that the sum of the residual squares became as minuscule as possible and then the least square solution for R_j was:

$$\frac{\partial\left(V_j' V_j\right)}{\partial R_j} = 0 \qquad (2.8)$$

$$R_j = \left(K'K\right)^{-1} K'S_j \qquad (2.9)$$

The three components' endmembers were then calculated after applying equation (2.9) to every spectral band in the QuickBird image of the sampled *An. arabiensis* habitat in the dataset.

We used UNIMAX to create fraction images of a prolific *An. arabiensis* habitat images. The unmixing algorithms were based on the following model, which assumed that a spectrum is a linear superposition of endmembers:

$$R_k = \sum_{i}^{n} a_i \cdot E_{i,k} + \varepsilon_k \qquad RMSE = \sqrt{\left(\sum_{k}^{m} \varepsilon_k^2\right)^{-m}}$$

R_k Reflectance of source wavelength k

$E_{k,i}$ Reflectance of endmember i at the wavelength k

a_i Abundance of endmember i

ε_k Error at wavelength k

RMSE Root mean square error of the ε_k

n Number of endmembers

m Number of wavelengths in the discrete spectrum

All algorithms in this research were given in nanometers (BEAM's default wavelength unit). The input signatures were assumed to represent spectral classes. The output fraction images stored values indicating the percentage of each sub-pixel estimate that composed the paddy preparation *An. arabiensis* habitat. UNMIX separated every habitat sub-pixel composition. UNMIX inputs observations of particulate pixel composition and seeks to find the number, composition, and contributions of the endmember data contributing sources or source types [2]. In this research UNMIX was also used to produce estimates of the uncertainties in the spectral predictors. UNMIX inputted the field and remote-sampled data in tabular format as flat ASCII files. The parameters used in UNMIX in this research was controlled by the following global parameters:

DBEM Endmember Signature Segment List

DBOC Database Output Channel List

RMSCHAN RMS-Error Output Channel

DBIW Database Input Window

RANGE Min and Max Digital Numbers

NORM Normalization: YES/NO

REPORT Report Mode: Term/Off/Filename

We specified the name of the database which contained the QuickBird image channels and endmember signatures of the sampled paddy preparation *An. arabiensis* habitat as EASI>FILE="QBfilespec." We then specified the segment numbers to be used as endmember signatures. The *An. arabiensis* habitat reference signature was then created using EASI>DBEM=*i,j,k*. The ranges of segments were specified with negative values. Signature segments were specified. We specified the output channels to receive the spectral endmember fractionalized data. There was one output channel for the input signature segment (EASI>DBOC=*i,j,k*). RMSCHAN was then used to specify the output channel to receive the residual error image as EASI>RMSCHAN=*i*. The DBIW specified the database input window to be unmixed as EASI>DBIW=Xoffset, Yoffset, Xsize, Y size. The RANGE was specified by the minimum and maximum output values which were used for scaling the fractionalized data. Note, that since unscaled

fraction image values can be outside the range of 0 to 1, then it is possible for the scaled output channels to contain values outside the range specified by RANGE [2]. In this research the default values for RANGE was 0 and 255. Additionally, EASI>RANGE=min,max specified whether output channels were to be normalized or not.

UNMIX performed the unmixing given the set of paddy preparation *An. arabiensis* habitat class segments (DBEM) created by the program. The signature was created using a training site bitmap. The output fraction images (DBOC) contained values which represented the percentage that of pixel constituents in the *An. arabiensis* habitat. The RMS (root mean squared) error was saved as RMSCHAN.

Before scaling, the output fractionalized image contained values which represented percentages between 0.0 and 1.0. Ideally, the sum of all unscaled fraction images should add up to approximately 1.0 at each pixel location [66]. The RANGE parameter was then used to specify the desired scaled output values which correspond to the unscaled values of 0.0 and 1.0 respectively. The RANGE defaulted to 0 and 255.

By default, the output image of the *An. arabiensis* habitat was not normalized (NORM="NO"). If the last class signature specified by DBEM represented a shadow area to be removed from the data, then normalization (NORM="YES") removed the effect of shadow from the habitat signature before saving the output. It is very important in spectral unmixing that the shadow area be specified as the last signature segment [66]. In our analyses we focused on the two terms of $R(i,v) = K_g G + \frac{C}{A}\int\int_{A_c} \frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v} ds$. The first term described how the sunlit background proportion proceeded to a maximum point as viewing and illumination positions in the hemisphere coincided. The second term described how the sunlit *An. arabiensis* habitat surface components composed of the Lambertian facets, became maximally exposed to view at the hotspot, while those facets on tops became dominant at large viewing zenith angles. The hot spot correlation effect refers to the observed brightening which can occur when viewing a scene from the same direction as the solar illumination [8] which for predictive vector insect habitat modeling is commonly, noted in the visible and NIR spectral regions [61].

2.11. Ordinary Interpolation Analyses

Spatial linear predictors generated from the paddy preparation *An. arabiensis* habitat were then generated using an Ordinary kriged-based interpolator. The Ordinary kriging algorithm was used to generate predictive maps of all productive habitats in the QuickBird image. Ordinary kriging was selected to interpolate the value $Z(x_0)$, (i.e., an immature paddy preparation *An. arabiensis* habitat count value), $Z(x)$, at an unobserved habitat location x_0 from the field and remote-sampled explanatory covariates, and $z_i = Z(x_i)$, i = 1..., *n* at nearby habitat locations, x_1, xn. [67]. In this research, Ordinary kriging was computed as a linear unbiased estimator, $\hat{Z}(x_o)$ of $Z(x_0)$ based on a stochastic model of the spatial dependence quantified by the variogram $\gamma(x,y)$ or by expectation $\mu(x) = E[Z(x)]$ and the covariance function $c(x,y)$ of the random field. The kriging estimator was given by a linear combination of the algorithm $: \hat{Z}(x_o) = \sum_{i=1}^{n} w_i(x_o) Z(x_i)$ using the endmember dataset of $z_i = Z(x_i)$ with weights $w_i(x_o)$, i = 1,...,*n* chosen, such that the variance was calculated using: $\sigma_i^2(x_0) := Var(\hat{Z}(x_0 o) - Z(x)) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i(x_0) w_j(x_0) c(x_i, x_j) + Var(Z(x)) - 2\sum_{i=1}^{n} w_i(x_0) c(x_i, x_0)$ which was further minimized using: $E[\hat{Z}(x) - Z(x)] = \sum_{i=1}^{n} w_i(x_0)\mu(x_i) - \mu(x_0) = 0$. This spatial interpolation method has been used for generating semivariograms and global autocorrelation statistics, in ArcGIS® Geostatistical Analyst, for habitats of anopheline mosquito habitats in two urban environments in Kenya (62).

In this research, the dependent variable was the spectral endmembers of the sampled habitat, which was transformed to fulfill the diagnostic normality test prior to performing kriging. The kriging weights of Ordinary kriging were used to fulfill the unbiasedness condition in the spatial interpolation of the ecologically sampled datasets using $\sum_{i=1}^{n} \lambda_i = 1$ which was given by the Ordinary kriging equation system:

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ 1 \end{pmatrix} = \begin{pmatrix} \gamma(x_1,x_1) & \cdots & \gamma(x_1,x_n) & 1 \\ \vdots & \ddots & & \vdots \\ \gamma(x_n,x_1) & \cdots & \gamma(x_n,x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1,x^*) \\ \vdots \\ \gamma(x_n,x^*) \\ 1 \end{pmatrix}$$

The additional parameter μ was a Lagrange multiplier used in the minimization of the kriging error, $\sigma_k^2(x)$, to maintain the unbiasedness condition in the ecological dataset We considered the optimization problem using maximize *f(x ,y)* subject to *g(x,y)*=c. We introduced a new variable (λ) (i.e., Lagrange multiplier) where the function was defined by $\Lambda(x,y,\lambda) = f(x,y) + \lambda \cdot (g(x,y) - c)$.

The λ term was added. If (*x*, *y*) is a maximum for the original constrained problem, then there exists a λ such that (*x*, *y*, λ) is a stationary point for the Lagrange function (stationary points are those points where the partial derivatives of Λ are zero) [49]. The Ordinary kriging was given by:

$$\hat{Z}(x^*) = \begin{pmatrix} z_1 \\ \vdots \\ z_n \\ 0 \end{pmatrix}' \begin{pmatrix} \gamma(x_1,x_1) & \cdots & \gamma(x_1,x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n,x_1) & \cdots & \gamma(x_n,x_n) & 1 \\ 0 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1,x^*) \\ \vdots \\ \gamma(x_n,x^*) \\ 1 \end{pmatrix}$$

The Ordinary kriging error was given by:

$$\mathrm{var}(\hat{Z}(x^*) - Z(x^*)) = \begin{pmatrix} \gamma(x_1,x^*) \\ \vdots \\ \gamma(x_n,x^*) \\ 1 \end{pmatrix} \begin{pmatrix} \gamma(x_1,x_1) & \cdots & \gamma(x_1,x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n,x_1) & \cdots & \gamma(x_n,x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1,x^*) \\ \vdots \\ \gamma(x_n,x^*) \\ 1 \end{pmatrix}$$

Conventional classification techniques require extensive ground truth information [68]. In this research, a confusion

Table 1. Information Collected in the Rice Fields of Karima Study Site for Analysis in SAS

Variable	Description	Units
Anopheles count	Total larval/pupal count (dependent variable)	Count
Tillers	Density	Number/Square meter
Depth	Field depth	Centimeters
Canopy	Canopy cover	Percent
Turbidity	Turbidity status	0 = not turbid, 1 = turbid
Disanimal	Distance to animal	Meters

Table 2. The Mean Number of *An. Arabiensis* Larvae Collected (Mean ± SE) Per 20 Dips in Paddies Containing Different Stages of Rice Growth Using Field Sampled and QuickBird 0.61 m Visible and Near Infra-red (NIR) data in Karima Rice-Village Complex

Paddy category	Number of habitats	1st instars	2nd unstars	3rd instars	4th instars	Pupae
Paddy preparation	30	8.02	9.00	8.89	3.22	0.99
Ploughed	25	1.41	0.95	0.09	0.00	0.36
Flooded	23	1.67	1.10	0.30	0.00	0.36
Tillering	28	2.00	6.67	2.00	1.20	0.67
Flowering/maturation	27	0.01	0.00	0,02	0.01	0.00
Fallow	27	1.00	0.67	0.00	0.00	0.01

matrix was calculated by comparing the predicting *An. arabiensis* habitat location and class of each ground truth pixel with the corresponding location and class in the predictive spatial linear model. Each column of the confusion matrix represented a ground truth class and the values in the column corresponded to the classification image's labeling of the ground truth pixels. Fortunately, the spectral characteristics of an unknown pixel in a Gaussian probability distribution using the spectral signature of the training data work on a pixel support or spatial resolution using classified land cover, which limits the amount of validation required to calibrate the final model [68]. The overall accuracy of the predictive estimates generated from the spectral decomposition of the QuickBird visible and NIR data of a paddy preparation *An. arabiensis* habitat was calculated by summing the number of pixels classified correctly and dividing by the total number of pixels. The pixels classified correctly were found along the diagonal of the confusion matrix table which listed the number of pixels that were classified into the correct ground truth class.

RESULTS

The relative abundance of the riceland *An. arabiensis* habitats were identified by the digitized grid. The overall accuracy of the land cover classification from the QuickBird image was 0.94 ($P < 0.0001$) for the riceland study site. Generally, the frequency of confusion of the classification was low. Overall, instances of confusion were minimal and did not affect user's classification accuracy for the study sites. The user's accuracy ranged between 92 and 97%, with relatively low errors of commission (excesses), and the producer's accuracy ranged between 91 and 98 % for the LULC classification.

Table **1** lists the dependent and independent variables collected in the study site. The count of *An. arabiensis* larvae/pupae collected at a habitat has a mean of 4.7, with a standard deviation of 5.6. The median count was 5 *Anopheles* larvae/pupae and the count range from 0 to 41 *Anopheles* larvae. The distribution was right skewed with 75% of the habitats having 4 or less larvae. Significant correlations exist between some of the independent variables including study site and number of tillers, (r=-0.23, p=0.004).

Table **2** revealed the abundance of riceland *An. arabiensis* larvae/20 dips collected in the habitats at the Karima study site. In the study site, the difference in the abundance of pupae and 1st, 2nd and 3rd instars larvae collected in paddy preparation and other anopheline habitats was not significant ($P > 0.05$). The 4th instars larvae; however, was significantly higher in the paddy preparation habitats than in the other habitats. ($t = 5.19$, df 179, $P < 0.05$).

The regression analyses identified depth of habitat as significantly influencing the count of *An. arabiensis*

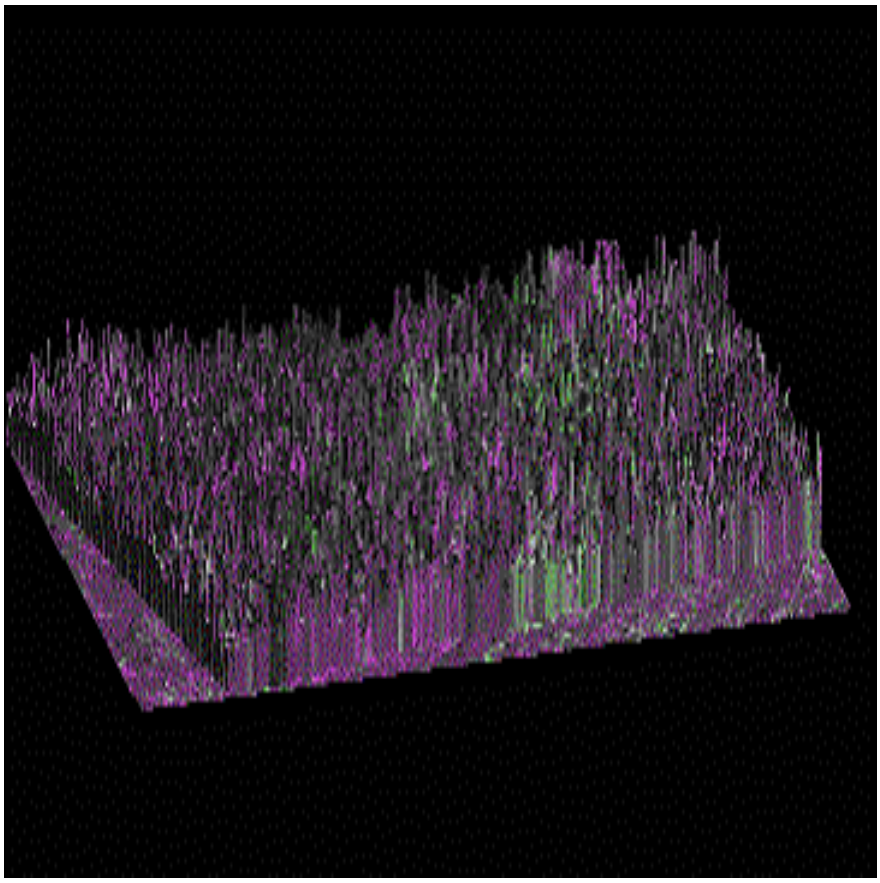

Fig. (3). A 3-D paddy preparation *An. arabiensis* habitat sampled in ENVI®.

mosquitoes in the study site. Mwangangi *et al*. [69] used a backward linear regression model and determined that abundance of *An. arabiensis* larvae at any of the rice growing phases included water depth, in paddy preparation habitats in the Mwea rice fields. By specifying coefficient estimates in a Bayesian conditional heteroskedastic framework, Jacob *et al*. [1] found depth of habitat was found to be a significant predictor positively associated with anopheline habitats.

The Feature Extraction Module in ENVI® allowed quick and accurate extraction of a paddy preparation *An. arabiensis* habitat from the QuickBird visible and NIR data (Fig. **3**). The library was resampled by averaging intermediate values to match QuickBird bandwidths; values between two bands were considered as zero during averaging original and resampled spectra with the full width at half maximum (FWHM) for each of the bands. The FWHM is a measure of the width of the line at a point that is half the line's peak value in a either an emission or absorption spectrum [62]. The paddy preparation *An. arabiensis* habitat pixel was then isolated from the LULC map.

The endmembers in the paddy preparation *An. arabiensis* habitat then was used to estimate the set of distinct spectra that comprised the mixed QuickBird pixel in the riceland scene. The inversion stage produced abundance planes that provided estimates of the fractional abundances for the endmembers in the decomposed pixel. An exact overlap function was also generated where $\theta_v' = \theta_i'$ and ϕ varied from 0 to 2π on the principal cone using a hybrid of geometric optical model for remotely capturing bidirectional reflectance over discontinuous plant riceland *An. arabiensis* canopies. According to our model, mutual shadowing of illumination did d not change the ratio $K_C/\left(1-K_g\right)$. This ratio was itself denoted K_c which we used to generate u A_C/A for determining consistency with K_g where the mutual shadowing in illumination and viewing directions was independent of A_c/A for consistency with K_g

In the direction of illumination, the *An. arabiensis* habitat had an area $\pi R^2 \sec\theta_i'$, and the total projected area was then calculated to be $\lambda\pi R^2 \sec\theta_i'$, if there was no mutual shadowing. Because of mutual shadowing, however, the net projected area was $1-e^{\lambda\pi R^2 \sec\theta_i'}$. We defined the quantity M_i, the mutual shadowing proportion in the illumination direction as $M_i = 1-\frac{1-e^{-\lambda\pi R^2 \sec\theta_i'}}{\lambda\pi R^2 \sec\theta_i'}$. M_i which revealed the degree of mutual shadowing in the illumination direction. In other words, each spheroid, on average, had a proportion M_i of the imaged *An. arabiensis* habitat surface

area that was not sunlit. This part of the habitat was concentrated at the lower part of the spheroid. We then generated a boundary drawn on the habitat surface of the spheroid with the area comprising M_i located below it. Similarly, we defined M_v as the mutual shadowing proportion of the rice plant components in the view direction as $M_v = 1 - \frac{1 - e^{-\lambda \pi R^2 \sec\theta_v'}}{\lambda \pi R^2 \sec\theta_v'}$.The viewing shadows were also concentrated at the lower part of the spheroid so we were able to define the M_v boundary. The proportion of sunlit the QuickBird sensor captured corresponded to the area above both M_i and M_v boundaries which depended on both zenith and azimuth differences between the illumination and view directions. At the hotspot, M_i and M_v boundaries overlapped and the data revealed no mutual shadowing of the *An. arabiensis* habitat components. In this research when the view zenith angle was larger than the illumination zenith angle, M_v was greater than M_i and little or no mutually-shaded habitat area was visible, based on the azimuth differences between the imaged objects. Thus, were able to capture the essence of the mutual-shading effect of the *An. arabiensis* habitat canopy components.

We then quantified the f-Ratio of non-nadir viewed spheroids. First, we considered a single spheroid in the decomposed QuickBird pixel spectral data. For the spheroidal case, it is necessary to show whether the f-ratio is still independent of density, as in the case of the nadir-viewing cones [20]. From the view direction, the spheroid had a projected area $\Gamma_v = \pi R^2 \sec\theta_v'$; however, only the portion $\frac{1}{2}\left(1 + \langle i', v' \rangle\right)$ of that *An. arabiensis* habitat was sunlit. Similarly, the illumination shadow on the ground occupied the habitat area $\pi R^2 \sec\theta_i'$. The compound area of the habitat plus illumination shadow projected onto the background was $\Gamma = \pi R^2 \left[\sec\theta_i' + \sec\theta_v' - O(\theta_i', \theta_v', \phi)\right]$. We then were able to define the ratio F for the spheroidal and its associated attributes as $F = \frac{\Gamma_c}{\Gamma} = \frac{\frac{1}{2}\left(1 + \langle i', v' \rangle\right)\sec\theta_v'}{\sec\theta_i' + \sec\theta_v' - O(\theta_i', \theta_v', \phi)}$, where Γ_c was the sunlit area of the *An. arabiensis* habitat and we defined the corresponding ratio $f = \frac{K_c}{1 - K_g}$ for the endmember selection of these sub-pixel emissivities. In this research, n represented the shadow parameters generated from the decomposed QuickBird *An. arabiensis* habitat pixel. If there was no mutual shadowing, $f = F$. As n increased, however, mutual shadowing occurred, and as such $K_g = e^{-\lambda \pi R^2 \left[\sec\theta_i' + \sec\theta_v' - O(\theta_i, \theta_v, \phi)\right]}$. We then defined the mutual shadowing proportion M as $M = 1 - \frac{1 - K_g}{\lambda \Gamma}$, which was the fraction of total shadowing cast from the LULC components that fell onto the *An. arabiensis* habitat instead of the background. The sunlit and viewed habitat surface features was reduced by hiding either from viewing or from illumination.

We then quantified the f-ratio with mutual shadowing which was $f = \frac{n\Gamma_c - \Sigma\Delta_{Ac}}{A(1 - K_g)} = F\frac{1 - \Sigma\frac{\Delta_{Ac}}{n\Gamma_c}}{1 - M}$ where $\sum \Delta_{A_c}$ was the total decrement from $n\Gamma_c$ to A_c (i.e., the background-projected area of viewed sunlit An. arabiensis habitat surface). We expressed $\sum \Delta_{A_c}$ as three terms: a decrement due to mutual shading in the view direction plus a decrement due to mutual shading in the sun direction, minus those elements shaded in both directions using $\sum \Delta_{A_c} = n\Gamma_v \left(P_v M_v + P_i M_i - P_o\right)$,where P_v was the conditional probability that the *An. arabiensis* habitat faced the sun given that it was mutually shaded from view. In this research P_i was the probability that the *An. arabiensis* habitat surface elements faced the viewer given that it was mutually shaded from illumination. Both P_i and P_v were average proportions of the habitat areas projected in the view direction. P_o, the third term, was the overlapped part of the first two terms, expressed as a fraction of Γ_v. In this research, P_o contained three parts derived from the habitat surface element and the rice- vegetation canopy structure. This collection contributed to the hotspot, due to the spatial correlation of the shadows. Since the probabilities of being hidden in multiple directions were not independent, we were able to substitute $\sum \Delta_{A_c} = n\Gamma_v \left(P_v M_v + P_i M_i - P_o\right)$ into $f = \frac{n\Gamma_c - \Sigma\Delta_{Ac}}{A(1 - K_g)} = F\frac{1 - \Sigma\frac{\Delta_{Ac}}{n\Gamma_c}}{1 - M}$ which yielded a single expression for $f = F\frac{1 - \Gamma_v \left(P_v M_v + P_i M_i - P_o\right)/\Gamma_c}{1 - M}$.

We then quantified P_v, P_i and P_o. We used all illumination or viewing shadows incorporating M_i or M_v boundaries respectively. In our model, P_v, P_i and P_o, where used to visualize the M_v and M_i boundaries. If viewing and illumination shadows fall strictly below M_v and M_i boundaries, then P_v, the conditional probability that at surface element facing the sun given the mutually shadowed areas will be the ratio of the illuminated portion of the projected surface below the M_v boundary [20]. Correspondingly, in this research, P_i was the conditional probability that the *An. arabiensis* habitat directly faced the viewer given that it was mutually shaded from illumination and was the ratio of the viewed portion of the projected habitat area below the M_i boundary. Note, that M_i was the proportion of mutually-shaded *An. arabiensis* habitat habitat surface projected to the direction of illumination, but

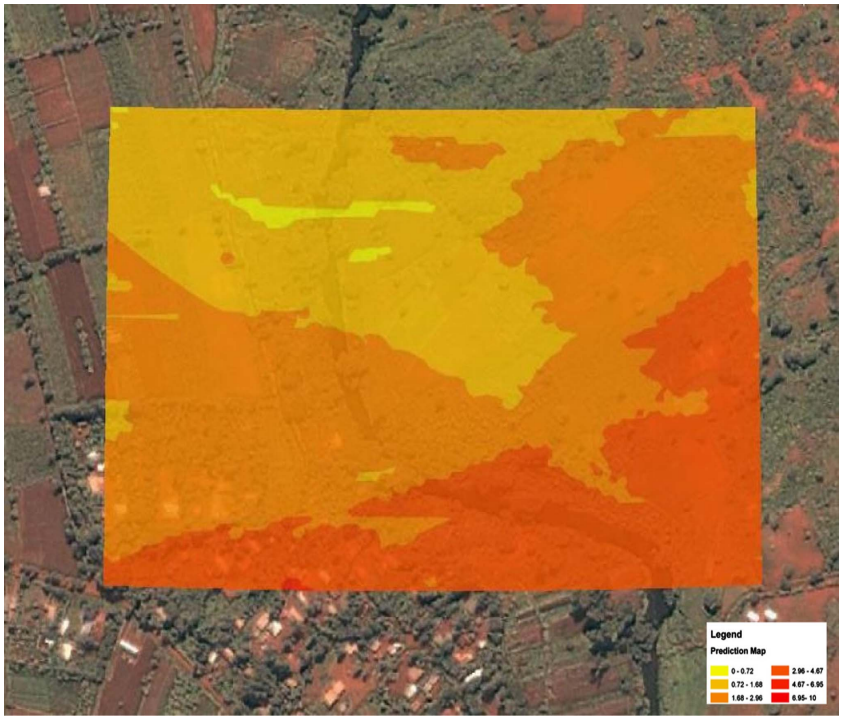

Fig. (4). Kriged larval/pupal abundance counts from the paddy preparation *An. arabiensis* habitat pixel spectral emissivity estimates sampled in the Karima study site

$P_i M_i \Gamma_v$ was the area of this fraction of the habitat surface which was projected to viewing direction. Proper calculation of this portion of the habitat and its associated attributes involved some projection change. We used P_o as the variable representing the overlap area o, which was also represented as a fraction of Γ_v. We then considered the case in the principal plane. For simplicity, we assumed that all shadows from the *An. arabiensis* habitat and its LULC components fell below the boundaries M_v and M_i, which were the traces of planes intersecting the spheroid at its center. The angle between the planes of the M_i and the illumination boundary was: $\theta_{M_i} = \cos^{-1}(1-2M_i)$. We then defined θ_{M_v} similarly. Our results indicate that in *An. arabiensis* habitat hotspot when the M_i and M_v boundaries coincided then $P_v = P_i = 1$, $P_o = M_v = M$, and $f = F = 1$. We then assumed that the viewing zenith angle increased to $\theta_v > \theta_i$. In usual cases when mutual shadowing of a sunlit spatial object , the M_v boundary is higher than the M_i boundary [5]. In the QuickBird sensor's view, P_v was the ratio of the *An. arabiensis* habitat's surface area between M_v boundary and the illumination boundary to the whole area under the M_v boundary. That is, $P_v = \dfrac{M_v\Gamma_v - (\Gamma_v - \Gamma_c)}{M_v\Gamma_v}$ while P_i was one, and P_o cancelled the M_i term. Then, the equation become $f = F\dfrac{1-P_vM_v\Gamma_v/\Gamma_c}{1-M} = F\dfrac{(1-M_v)\Gamma_v}{(1-M)\Gamma_c} = \dfrac{1-e^{-\lambda\Gamma_v}}{(1-K_g)}$. This result suggested that when the viewing direction in the principal plane deviated from $(\theta_v > \theta_i)$, the f-ratio will change. When the coverage is very low, the increment of θ_{M_v} may be also so small that M_v will be under the M_i boundary [2]. In this research we found that when θ_v moved inward on the principal plane but had not reached nadir, the M_i was higher than M_v: hence, $P_v = 1$, $P_o = M_v$, and $P_i = \dfrac{1-\cos(\theta_{M_i} - \theta_i' + \theta_v'\cos\phi)}{1-\cos\theta_{M_i}}$ After θ_v passed the nadir, the M_v boundary went to the opposite side of the spheroid from M_i. In this case, the QuickBird sub-pixel spectral data revealed the horizontal projection of the habitat and its LULC components at $\phi = \pi/2$. We then used P_i with ϕ equal to π and P_v which was the fraction of M_v over the illumination boundary which rendered:

$$P_v = \begin{cases} \dfrac{1-\cos(\theta_{M_v} - \theta_v' + \theta_v'\cos\theta)}{1-\cos\theta_{M_v}}, & (\theta_{M_v} - \theta_v' + \theta_v'\cos\theta) \geq 0 \\ 0, & (\theta_{M_v} - \theta_v' + \theta_v'\cos\theta) < 0 \end{cases}$$

Our residual model estimates also revealed that when θ_v was between the *An. arabiensis* habitat hotspot and nadir, P_v was always one, and, thus, a discontinuity of P_v appeared at the nadir. This discontinuity arose from the

assumption that all shadows fell under the M_v boundary. Additionally, in this research, the M_v at $\theta_v = 0$ was the physical intersection of boundaries between the *An. arabiensis* habitat and its LULC components, which did not change with viewing geometry, thus, $P_v M_v$ was still continuous at nadir, and equal to P_o. In other words, the formula had a very large viewing zenith, so that $\theta_{M_v} - \theta_i' + \theta_v' \cos\phi) > 0$.

In this research when M_i and M_v were independent, $\sum \Delta_{A_c} / (n\Gamma_c) = M$,. $f = \frac{n\Gamma_C - \Sigma\Delta_{AC}}{A(1-K_g)} = F\frac{1-\Sigma\frac{\Delta_{AC}}{n\Gamma_C}}{1-M}$. If all the spatial objects are at the same height, the situation will be very close to the "uniform height case" – mutual shadows will always fall on the lower part of the objects and the object top-viewing effect will be strong; however, when heights are distributed over a wide range, the top layer of the canopy will play a more important role in determining the BRDF of the canopy than the lower layer [20]. Therefore, when the *An. arabiensis* habitat structural heights were quantified by spectral distribution, the BRDF was apparent. The BRDF was determined by the size, shape, and height of the habitat and its associated LULC components in the top layer. Thus, we restricted ourselves to considering a single top layer only, where the range of distribution of height of the sampled habitat and its attributes did not exceed twice the vertical axis of the spheroid. To share the weighting between the spatiotemporal-sampled explanatory spectrally dependent covariates we used the parameter $\beta = \left(1 - \frac{h_2 - h_1}{4b}\right)^2$ Then both P_v and P_i were calculated as a weighted sum of corresponding terms $P = \beta P_1 + (1-\beta)P_2$, where P_1 and P_2 were the spectral probabilities associated with the spatial dimensions of the sampled habitat. The resulting scene was broken down into their canopy fractions specifically sunlit and shadowed background and scene brightness. Illumination direction was calculated by a linear combination of the canopy fractions and their respective radiance estimates.

An expression for additional azimuthual variation was also derived from the geometric-optical model. This azimuthual variation differed fundamentally in radiance for each elementary layer of the *An. arabiensis* habitat canopy. It was observed that all non-zero polar angles were most evident in the canopy when vertical and nearly opaque components of the habitat and its LULC components which were illuminated and viewed along polar sun angles. For the variation of the directional reflectance of the *An. arabiensis* habitat canopy cover with azimuthual view angle, shade-related parameters were quantified when the illuminated area of the imaged habitat (i.e., areas that was affected by the sun at large angles from the zenith) was remotely identified. Our results also indicated that the cause of the azimuthual variation could be traced to solar flux illumination of the vertically-oriented rice plant components and the variation of reflectance moderated by azimuthally isotropic sources of flux from sky light and the habitat canopy reflectance values. A scattergram representing the endmember reference signature of the *An. arabiensis* habitat spectral reflectance values was then generated.

Semivariogram plot of the logit scale model residuals was constructed, confirming a short range spatial pattern up to a distance of about 5 km. To carry out this process, residuals for all observed points were calculated on the logit ($\ln(p/1-p)$) scale of the model. The kriging module included a *An. arabiensis* habitat variogram model represented as :

$$\tilde{\gamma}(h) = \begin{cases} C_0 + C_1\left(1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right) & if\ |h| \le a \\ C_0 + C_1 & if\ |h| > 0 \end{cases}$$

A kriged map of deviance residuals was then calculated which was added to the predicted values on the logit scale before transforming the results back to proportions for the study site. Spatial dependence, displayed by these plots, was modeled using a constructed semivariogram. The addition of kriged residuals allowed the output to deviate from the model; the deviation was supported by the endmember spectral data. These smoothed values improved the final maps of the paddy preparation *An. arabiensis* habitat data sampled in the study site. An exponential model was fitted to the semivariogram, using a range of 72.6 m, a nugget of 0.21 (variance), a lag size of 11.5 m with 12 lags and a partial sill of 0.24 (variance) for the riceland study site (Fig. **4**).

We then calculated the kriged-based interpolation error. In this research, the error of *i*-th estimate, ri, was the difference of estimated value and true value at that sampled *An. arabiensis* habitat location: $r_i = \hat{v} - v_i$. The average error of a set of k estimates was: $m_r = \frac{1}{k}\sum_{i=1}^{k} r_i = \frac{1}{k}\sum_{i=1}^{k}\hat{v}_i - v_i$, and the model autoregressive error variance was: $\sigma_R^2 = \frac{1}{k}\sum_{i=1}^{k}(r_i - m_R)^2 = \frac{1}{k}\sum_{i=1}^{k}\left[\hat{v}_i - v_i - \frac{1}{k}\sum_{i=1}^{k}(\hat{v}_i - v_i)\right]^2$ In this research V(Xi). Xi was the habitat location of observed data for i > 0 and i <= n. The unknown value at the predicted habitat location was V(X0). The final kriged-based interpolation error variance was $\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \tilde{C}_{ij} - 2\sum_{i=1}^{n} w_i \tilde{C}_{i0} + 2\mu\left(\sum_{i=1}^{n} w_i - 1\right)$ In order to get the minimum variance of error, we calculated the partial first derivatives of the equation for each w and setting the result to 0. Herewith the differentiation with respect to w1 was $\frac{\partial(\sigma_R^2)}{\partial w_1} = 2\sum_{j=1}^{n} w_j \tilde{C}_{1j} - 2\tilde{C}_{10} + 2\mu = 0 \qquad \sum_{j=1}^{n} w_j \tilde{C}_{1j} + \mu = \tilde{C}_{10}$. All of weight Wi in the model was represented as: $\sum_{j=1}^{n} w_j \tilde{C}_{ij} + \mu = \tilde{C}_{i0}$ for each i, 1 <= i <= n**.** We quantified each weight Wi through the equation. After generating the value, we estimated the habitat value located in X0. In this research we used the *An. arabiensis* habitat variogram instead of covariance to calculate each weight of the equation. The final variogram was $\gamma_{ij} = \tilde{\delta}^2 - \tilde{C}_{ij}$ and the minimized estimation variance was $\tilde{\delta_R}^2 = \sum_{i=1}^{n} w_i \gamma_{i0} + \mu$.

DISCUSSION

The regression analyses identified paddy preparation as the rice-stage with the highest larval/pupal counts. The occurrence of higher immature *An. arabiensis* counts during the paddy preparation stage of rice cycle has been attributed to the presence of numerous open sun lit pools created by rice workers [70-72]. These results suggest that agricultural practices have significant influence on mosquito species diversity and abundance and that certain habitat characteristics favor production of malaria vectors. These factors should be considered when implementing larval control strategies which should be targeted based on habitat productivity and water management.

In this research, the residuals from the regression were the spatiotemporal-sampled *An. arabiensis* habitat values of the dependent variable Y minus the estimates of their values and the independent variable X. Since the i^{th} datum was (x_i, y_i) and the equation of the regression line was $y = a \times x + b$, the i^{th} residual was $e_i = y_i - (a \times x_i + b)$. A residual plot (i.e. scatterplot of the n points (x_i, e_i), i = 1, … , n) was then generated. Scatterplots shows heteroscedasticity, nonlinear association, or outliers if and only if the original scatterplot does, but it is easier to see these qualitative features of bivariate data in the residual plot than in the scatterplot of the original data [20]. We found that the correlation coefficient between the residuals and the independent variable was zero—the residuals did not have a trend with X—and the average of the residuals was zero. If the residuals have a trend in a linear predictive vector mosquito habitat model, the slope of the regression line is computed incorrectly but, if the residuals do not have a trend and the mean of the residuals is not zero then the intercept of the regression line is computed incorrectly[1]. In this research, the rms of the residuals, measured the average error of the regression line while estimating the dependent variable Y from the independent variable X. The rms error of regression was dependent only on the correlation coefficient of X and Y and the SD of Y: rms error of regression = $(1 - (r_{XY})^2)^{½} \times SD_Y$.

In this research, ENVI® software automatically categorized individual pixels of paddy preparation *An. arabiensis* habitats into separate spectral classes, converted remotely sensed raster layers to vector coverages and classified the layers as ESRI shapefiles. ENVI® data displayed all the georeferenced habitats by LULC data. The object based classifier provided methods for locating specific paddy preparation *An. arabiensis* pixels for interactive spatial/spectral pixel editing. For example, the ENVI® Feature Extraction module helped automate the process of performing accurate segmentations for image classification of the paddy preparation habitat. Additionally, the FLAASH module in ENVI® provided an accurate means of compensating for atmospheric effects in the riceland environments. Most atmospheric correction programs do not consider properties such as elevation, water vapor, and aerosol distribution [73]. The FLAASH model included a method for retrieving an estimated aerosol/haze amount from selected "dark" land covers pixels of the paddy preparation *An. arabiensis* habitats in the QuickBird scene. Use of ENVI® for ArcGIS® supported the production of intuitive workflows in the ENVI® module for conducting tour object-oriented classification using the field and remote-sampled riceland habitat data.

Independently determined spectral signatures for each of shadowed or sunlit habitat components and other sampled covariates were then calculated by the Li-Strahler geometric–optical model using the spectral bidirectional reflectance of the habitat surface variables. The model quantified predictor variables such as rice vegetation canopy as an assemblage of partially illuminated spatial objects using geometric optics and Boolean set theory. Our model also established the proportion of shadowed canopy and background of the sampled paddy preparation *An. arabiensis* habitat pixel and their associated variables as a function of view angle and illumination angle. Illumination angles determine the primary path that incident light will trace into the rice-canopy and control the relative proportion of plant tissue that will be illuminated [70]. Directional reflectance of the geometric-optical model was then integrated to generate hemispherical reflectance for given illumination angles, using the Boolean algorithms, canopy characteristics and spectral signatures generated from the paddy preparation *An. arabiensis* habitat.

One of the main input variables of the inverted model was the fractional component of sunlit background. The model also removed the effect of shadow from the paddy preparation *An. arabiensis* habitat signatures before saving the output fractional images. Samples measured directly in the field are often disturbed in some way, removing the influence of the components' spatial distribution (leaf angle distribution, clumping index, etc.) and canopy multiple scattering contributions based on spectral intensity which are especially important to shaded endmembers [74]. Apparent spectral reflectance characteristic shaded components are often ignored, or grouped together to form one endmember, regardless of the spectral differences between each endmember [75]. In this research the understory shading caused by solar zenith angle variation in sampled riceland *An. arabiensis* habitats was quantified by the geometric-optical model. Geometric-optical models can be used to estimate sunlit canopy component reflectance and shadow reflectance for predicting biophysical parameters at sub-pixel scales [76]. The shadow area of a sampled paddy preparation *An. arabiensis* habitat was specified as a separate spectral signature segment.

The habitat pixel spectral signatures were the validated using the linear umixing algorithm but did not include any treatment of diffuse irradiance and tiller specularity. This required the selection of spectral signatures for each endmember which was calculated by the unmixing algorithm. The unmixing algorithm decomposed the paddy preparation *An. arabiensis* habitat into their respective endmember and abundances. The model revealed that total surface area of the paddy preparation habitat pixel was divided proportionally according to the fractional abundances of the *An. arabiensis* habitat surface

components. The endmember validation model revealed that the reflected radiation of a sampled habitat conveyed the correct spectral proportions of habitat characteristics and their associated surface materials. In this sense, a linear relationship existed between the fractional abundance of the substances comprising the paddy preparation habitat surface area being imaged and the spectrum of the reflected *An. arabiensis* habitat radiance estimates. The unmixing classification used the reflected spectral data for separating significantly contrasted pixels (e.g., a vegetated pixel next to a water pixel) and also separated regions within a paddy preparation habitat that were not excessively different based on pixel digital values (e.g., rice plant pixel next to a grass pixel).

Although endmember extraction algorithms can determine factors that contribute to producing the digital number, the influence of the local topography on reflectance has be corrected prior to pixel spectral reflectance estimation [66]. For example, our models revealed that the digital value of a QuickBird *An. arabiensis* habitat pixel corresponded to geometric and radiometric parameters of the calculated radiance. However, where as the path radiance and the upward transmittance dependent on the thickness of the atmosphere between the habitat surface and the QuickBird sensor (i.e. the elevation of the target), only specific components were modified by the relative orientation of the target pixel. Thus, the direct and the diffuse irradiance captured of the paddy preparation *An. arabiensis* habitat pixel may have been dependent on the angle between the habitat surface and the sun ray direction. As such, in this research a direct irradiant was computed by means of the cosine law from slope and aspects of the paddy preparation *An. arabiensis* habitat. A mask of cast shadows at the time of image acquisition was also computed to isolate habitat areas which were not getting any direct irradiance. To estimate the environmental irradiance for the riceland study site, a terrain view factor was computed in an analogous way. We were then able to account for the irradiance of the reflecting habitat surface components while encompassing the distance and atmospheric transmittance. Finally, the DN's in the image were converted to scene radiances by means of the calibration coefficients provided by the paddy preparation *An. arabiensis* habitat pixel surface reflectance.

Semivariograms were then generated in ArcGIS® Spatial Analyst which modeled the structure of spectral variability in the extracted paddy preparation habitat pixel. The semiovariogram was useful for summarizing the spatial continuity of the prolific habitat. Optimal linear predictors were generated by incorporating a model of the covariance of the random function using a weighted moving average interpolation and the QuickBird pixel spectral estimates. The model measured the local variation in the paddy preparation *An. arabiensis* sampled data as a function of distance and direction. The kriged smoothed maps, generated through the regression models displayed spatial patterns of the habitats. The kriged models provided a method for not constraining the An. arabiensis reference signature. In turn, the kriging algorithm allowed distance and direction in the interpolation process to be analyzed, which minimized the variance of unexpected error.

One of the advantages of quantifying pixel reflectance heterogeneity at the habitat scale using variogram models is statistical measures generated may also be parameters in analytical expressions that represent all, or part of a probability density function. Techniques vested with the assumption that the data originate from a parameterized probability density function are considered parametric [64]. In this research, polygons were digitized based on the georeferenced images within a GIS, which helped build the models based on the field and remote-sampled data for geostatistical interpolation of the *An. arabiensis* maps. Kriging was applied to the centroids of the polygons/digitized grid cells for calculating the maps, based on the pixel spectral reflectance estimates. The quality assessment of the maps was conducted by using statistical mean values. Predictive parametric algorithms generated using sub-pixel radiance estimates extracted from a paddy preparation *An. arabiensis* habitat can develop powerful predictor models for identifying other prolific habitats while incorporating many complexities (e.g. clustering of habitats based on larval/pupal productivity).

The pioneering nature of the conceptualization and analysis presented in this research alludes to many themes meriting future endmember *An. arabiensis* habitat research For example, spectral unmixing may be applicable to hyperspectral imagery and QuickBird data for mapping vegetation parameters of georeferenced spatiotemporal-sampled paddy preparation *An. arabiensis* habitats. In previous research hyperspectral bidirectional reflectance distribution function data of Konza prairie grassland acquired in the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE) on the ground captured with two SE-590 instruments and with the airborne advanced solid-state array spectroradiometer (ASAS) were analyzed and compared to BRDF data of dense ryegrass obtained with the European goniometric facility (EGO) and the Swiss field-goniometer system (FIGOS)[20]. The soil underlying the relatively sparse Konza prairie grass disturbed the spectral BRDF effects of the vegetation components. After a correction of the soil influence based on the bidirectional canopy gap probability, the Konza data from SE-590 and ASAS sensors revealed a consistently strong dependence of spectral BRDF effects from nadir reflectance as was observed in the EGO and FIGOS data. Indices such as Normalized Difference Vegetation Index (NDVI) Simple Ratio Index (SRI) and the Soil Adjusted vegetation Index (SAVI) thus may provide data information on vegetation-related parameters associated to productive *An. arabiensis* habitat using QuickBird and hyperspectral data. Although the potential of hyperspectral remote sensing and QuickBird data for spatially targeting *An. arabiensis* habitats is exciting there are special issues that arise with this unique type of imagery. For example, many of the hyperspectral analyses algorithms require accurate atmospheric corrections to be performed. Because the solar radiation on the sun-surface-sensor path in the 0.4–2.5 μm visible and NIR spectral regions is subject to absorption and

scattering by atmospheric gases and aerosols, the hyperspectral imaging data contains atmospheric effects [5]. Therefore, In order to use hyperspectral imaging data for quantitative remote sensing of land surfaces for identifying *An. arabiensis* habitats the atmospheric effects must be removed. Fortunately, over the years, atmospheric correction algorithms have evolved from the earlier empirical line method and the flat field method to more recent methods based on rigorous radiative transfer modeling approaches. Sophisticated atmospheric corrections algorithms have also been developed to calculate concentrations of atmospheric gases directly from the detailed spectral information contained in the imagery itself without ancillary data. These corrections can be performed separately for each pixel. Several of these atmospheric correction algorithms are available within commercial image processing software. Unfortunately, there are only few hyperspectral sensors presently available. Nevertheless, hyperspectral sensors are acquiring imagery from space including the Hyerion sensor on NASA's EO-1 satellite, the CHRIS sensor on the European Space Agency's PROBA satellite and the FTHSI sensor on the US Air Force Research Labs. Mighty Sat II satellite and The EROS Data Center currently provides Hyperion imagery at a relatively low cost to the general public. Many airborne hyperspectral sensor including NASA's AVIRIS sensor are also available to collect data. Spectral indices obtained from ground-based hyperspectral instruments such as spectroradiometer along with QuickBird data may then be a used to quantify vegetation covaraites associated to productive *An. arabiensis* habitats.

In the future we may also statistically classify riceland *An. arabiensis* habitats varieties using canopy BRDF data throughout the rice cycle. Mwangangi *et al.* [69] used an experimental paddy field to determine environmental conditions such as soil, nutrients, predators and other *An. arabiensis* habitat predictor variables were not homogeneous throughout the rice cycle. Spectral reflectance of each of riceland habitat varieties should be measured at nadir and at off-nadir angles of 45°, 30°, 15°, –15°, –30°, and –45° on both the principal and perpendicular planes at intervals of 1 nm from 400 to 850 nm, for example. The reflectances in QuickBird visible and NIR bands at every measuring angle may then be computed for each riceland *An. arabiensis* habitat variety. As a result the number of riceland habitat varieties that can be statistically distinguished using BRDF data may be larger than the number that can be distinguished using QuickBird spectral reflectance data at the nadir angle. The difference in BRDF among riceland *An. arabiensis* habitat varieties may be statistically significant.

In conclusion, a regression model using QuickBird LULC covariates and spatiotemporal field-sampled data revealed significantly higher *An. arabiensis* larval/pupal counts during the paddy preparation stage of rice development. ENVI® categorized individual pixels of a paddy preparation *An. arabiensis* habitats into separate spectral classes, converted remotely sensed raster layers to vector coverages and classified the layers as ESRI shapefiles. Topographic effects on the bidirectional and hemispherical reflectances generated from the habitat surface were then calculated with a geometric-optical model. Results indicated that the empirical and cosine correction used on the decomposed data was effective for all QuickBird spectral bands in both the solar and view directions. Our model.used a hemispherical integration of the BRDF which also provided an albedo estimate for the heterogeneous topographic habitat surface. Mutual shadowing proportions generated from the *An. arabiensis* habitat and its associated components were also quantified**.** Additionally, the model simulated accurately the changing of "hotspot" location with increasing view zenith angles in different areas of the sampled habitat based on the different solar zenith angles. The final model could simulate the angular distribution characteristic of reflectance spectrum of riceland *An. arabiensis* habitat canopies and the unsymmetrical distribution of the hotspot effect well. The model also revealed that the paddy preparation *An. arabiensis* habitats and their associated land cover attributes generated reflectance values based on areal proportions for the radiometric elements sunlit canopy, sunlit background and shadow fraction. The unmixing algorithm then identified the fractional presence of each endmember associated with the paddy preparation *An. arabiensis* habitats using separate sub-pixel spectral estimates. Thereafter, a kriged-based interpolation map displayed the spatial patterns of all other productive *An. arabiensis* habitats based on the reference signature. Spectrally quantifying QuickBird sub-pixel reflectance estimates from a productive *An. arabiensis* habitat can help guide larval control by spatially predicting other highly prolific unsampled habitats based on endmember spectral radiance.

ACKNOWLEDGEMENTS

This research was funded by the National Institute of Health Grant U01A154889 (Novak Robert) University of Alabama at Birmingham.

REFERENCES

[1] Jacob BG, Griffith D, Muturi E, *et al.* Describing Anopheles arabiensis aquatic habitats in two riceland agro-ecosystems in Mwea, Kenya using a negative binomial regression model with a non-homogenous mean. Act Trop 2009; 109(1): 17-26

[2] Jensen JR. Introductory digital image processing: A remote sensing perspective. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall 2005.

[3] Keshava N, Mustard JF. Spectral unmixing. IEEE Signal Process Mag 2002: 44-57.

[4] Woodcock C, Harward VJ. Nested-hierarchical scene models and image segmentation. Remote Sens Environ 1992; 25: 323-48.

[5] Burrough PA, Mc Donnell RA. Principles of geographical information systems. Oxford University Press: UK 1998.

[6] Gu W, Novak RJ. Habitat-based modeling of impacts of mosquito larval interventions on entomological inoculation rates, incidence, and prevalence of malaria. Am J Trop Med Hyg 2005; 73(3): 546-52.

[7] Foody GM. Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. Int J Rem Sens 1996; 17(17): 1317-40.

[8] Liu X, Skidmore AK, Van Oesten H. Integration of classification methods for improvement of land-cover map accuracy. ISPRS J Photogram Remote Sens 2002; 56(4): 257-68.

[9] Thomson AG, Fuller RM, Eastwood JA. Supervised versus unsupervised methods for classification of coasts and river orridors from airborne remote sensing. Int J Remote Sens 1998; 19(17): 3423-31.

[10] Pan YZ, Li XB, Gong P, *et al.* An integrative classification of vegetation in China based on NOAA AVHRR and vegetation-climate indices of the Holdridge life zone. Int J Remote Sens 2003; 24(5): 1009-27.

[11] Hu YH, Lee HB. Optimal Linear Spectral Unmixing. IEEE Trans Geosci Remote Sens 1993; 37(1): 639-44.

[12] Vikhamar D, Solberg R. Snow-cover mapping in forest by constrained linear spectral unmixing of MODIS data. Remote Sens Environ 2003; 8(33): 309-23.

[13] Zhu HL. Linear spectral unmixing assisted by probability guided and minimun residual exhaustive search for subpixel classification. Int J Remote Sens 2005; 26(24): 5585-601.

[14] Lu DS, Weng QH. Spectral mixture analysis of ASTER images for examining the relationship between urban thermal features and biophysical descriptors in Indianapolis, Indiana, USA. Remote Environ 2006; 104(2): 157-67.

[15] Bastin L. Comparison of fuzzy c-means classification, linear mixture modelling and MLC probabilities as tools for unmixing coarse pixels. Int J Remote Sens 1997; 18(17): 3629-48.

[16] Foody GM. Hard and soft classification by a neural network with non-exhaustively defined set of classes. Int J Remote Sens 2000; 23(18): 3853-64.

[17] Jacob BG, Muturi E, Halbig P, *et al.* Environmental abundance of *Anopheles* (Diptera: Culicidae) larval habitats on land cover change sites in Karima village, Mwea rice scheme, Kenya. Am J Trop Med Hyg 2007; 76(1): 73-80.

[18] Peddle DR, Smith AM. Spectral mixture analysis of agricultural crops: endmember validation and biophysical estimation in potato plots. Int J Rem Sens 2005; 26(22): 4959-79.

[19] Gao A simple bidirectional-reflectance model applied to tall grass canopy Remote Sens Environ 1993; 45 (2); 209-224.

[20] Schowengerdt RA. Remote Sens. 2nd ed. San Diego1997.

[21] Li J, Bruce LM, Byrd J, Barnett J. Automated detection of Pueraria montana (kudzu) through Haar analysis of hyperspectral reflectance data. Proc IEEE IGARSS2001; 5: 2247-9.

[22] Li J, Bruce LM. Improving the accuracy of linear pixel unmixing via appropriate endmember dimensionality reduction. Adv Techn Anal Remote Sensed Data 2003; 157-62.

[23] Hlavka CA, Spanner MA. Unmixing AVHRR imagery to assess clearcuts and forest regrowth in Oregon. IEEE Trans Geosci Remote Sens 1995; 33:788-95.

[24] Scholl JF, Dereniak EL. Hyperspectral wavelet transforms for data compression and feature extraction. Frontiers in Optics; Tucson, Arizona: Optical Society of America 2003.

[25] Scholl JF, Dereniak EL. Fast Wavelet based feature extraction of spatial and spectral information from hyperspectral datacubes. Proc SPIE 2004; 285-93.

[26] Sweldens W. The lifting scheme: A custom-design construction of biothogonal wavelets. Appl Comp Harm Anal 1996; 3(2): 186-200.

[27] Calderbank R, Daubechies I, Sweldens W, Yeo B-L. Wavelet transforms that map integers to integers. Appl Comp Harmonic Anal 1998; 5(3): 332-69.

[28] Barrett HH, Myers KJ. Foundations of Image Science. J Electron Imag 2005; 14(2); 37-44.

[29] Maselli F. Multiclass spectral decomposition of remotely sensed scenes by selective pixel unmixing. IEEE Trans Geosci Remote Sens 1998; 36: 1809-20.

[30] Luo J, King RL, Younan N. An unmixing algorithm based on vicinal information. Geosci Remote Sens Symposium 2002.

[31] Herries G, Seligo T, Danaher S. Singular value decomposition in applied remote sensing. IEE Colloquium Image Process Remote Sens1996; 5: 1-6.

[32] Herries G, Danaher S, Selige T. Characterization of agricultural land using singular value decomposition. SPIE 1995; 2585: 2-11.

[33] Sirkeci B, Brady D, Burman J. Restricted total least squared solutions for hyperspectral imagery. IEEE International Conference 2000; 1.

[34] Qian S. Hyperspectral data compression using a fast vector quantization algorithm. IEEE Trans Geosci Remote Sens 2004; 42(8): 1791-8.

[35] Petrou M, Foschi PG. Confidence in linear spectral unmixing of single pixels. IEEE Trans Geosci Remote Sens 1999; 37(1): 624-6.

[36] Nascimento JMP, Dias JMB. Does independent component analysis play a role in unmixing hyperspectral data? IEEE Trans Geosci Remote Sens 2005; 43(1): 175-87.

[37] Keshava N. Best band selection for detection in hyperspectral processing. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2001; 5: 3149-52.

[38] Zhu H, Eastman JR. Probability guided and minimum residual eshaustive searching approaches for subpixel classification. Proc IEEE Int Geosci Remote Sci 2001; 4: 1883-5.

[39] Ma JP, Zheng ZB, Tong QX. *et al.* An application of genetic algorithms on band selection for hyperspectral image classification. International Conference on Machine Learning and Cybernetics 2003; 5: 2810-3.

[40] Riedmann M, Milton EJ. Supervised band selection for optimal use of data from airborne syperspectral sensors. Proc IEEE IGARSS 2003; 3: 1770-2.

[41] Chein-I Chang, Du Q, Sun T-L, *et al.* A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. IEEE Trans Geosci Remote Sens 1999; 37(6): 2631-41.

[42] Haertel VF, Shimabukuro YE. Spectral linear mixing model in low spatial resolution image data. IEEE Trans Geosci Remote Sens 2005; 43(11): 2555-62.

[43] Zurita-Milla R, Clevers JGPW, Schaepman ME. Landsat TM, Meris FR. Image fusion for land cover mapping over the Netherlands. Proceedings of the 2nd Workshop of the EARSel SIG on Land Use and Land Cover 2006: pp. 34-40.

[44] Okomoto K, Fukuhara M. Estimation of paddy field area using the area ratio of categories in each mixel of Landsat TM. Int J Remote Sens 1996; 17(9): pp. 1735-49.

[45] Fang H. Rice Crop area estimation of an administrative division in China using remote sensing. Int J Remote Sens 1998; 19: 3411-9.

[46] Xiao X, Boles S, Frolking S, Salas W, *et al*. Landscape-scape characterization of cropland in China using Vegetation and Landsat TM images. Int J Remote Sens 2002; 23(18): 3579-94.

[47] Jacob BG, Arheart KL, Griffith DA, *et al.* Evaluation of environmental data for identification of Anopheles (Diptera: Culicidae) aquatic larval habitats in Kisumu and Malindi, Kenya. J Med Entomol 2005; 42(5): 751-5.

[48] Mwangangi JM, Muturi EJ, Shililu JI, *et al*. Distribution of mosquito larvae within the paddy and its implication in larvicidal application in Mwea rice irrigation scheme, Central Kenya. J Am Mosq Cont Ass 2008; 24(1): 36-41.

[49] Jacob BG, Muturi EJ, Mwangangi JM, *et al.* Remote and field level quantification of vegetation covariates for malaria mapping in three rice agro-village complexes in Central Kenya. Intl J Hlth Geog 2007; 6: 21-8.

[50] Service MW. Mosquito ecology: field sampling methods, 2nd ed. Essex, UK: Elsevier Publishers 1993.

[51] Edwards FW. Mosquitoes of the Ethiopian Region III. Culicine Adults and Pupae. London: British Museum (Natural History); 1941.

[52] Gillies MT, Coetzee M. A supplement to the Anophelinae of Africa south of the Sahara (Afrotropical region). Johannesburg, South Africa: South African Institute for Medical Research 1987.

[53] Rejmankova E, Roberts DR, Pawley A, *et al*. Predictions of adult *Anopheles albimanus* densities in villages based on distances to remotely sensed larval habitats. Am J Trop Med Hyg 1995; 53(5): 482-8.

[54] Beck LR, Rodriguez MH, Dister SW, *et al.* Remote sensing as a landscape epidemiologic tool to identify villages at high risk malaria transmission. Am J Trop Med Hyg 1994; 51(3): 271-80.

[55] Wood B, Washino R, Beck L, *et al.* Distinguishing high and low anopheline- producing rice fields using remote sensing and GIS technologies. Prev Vet Med 1991; 11: 277-88.

[56] Jacob BG, Muturi EJ, Funes JE, *et al.* A grid-based infrastructure for ecological forecasting of rice land *Anopheles arabiensis* aquatic larval habitats. Malaria J 2006; 5: 91.

[57] Dolo G. Etude des populations d' *An. gambiae* s.l. par marquage, lacher et recapture en 1993 et 1994 a Banambani (Arrondissement central de Kati, Mali). Memoire de DEA en Entomologie et Parasitologie medicales/ ISFRA1996.

[58] Stehman SV, Czaplewski RL. Design and analysis for thematic map accuracy assessment: fundamental principals. Remote Sens Environ 1998; 64(3): 331-44.

[59] Story M, Congalton RG. Accuracy assessment: a user's perspective. Photogram Eng Remote Sens 1986; 52(3): 397-9.

[60] Jacob BG, Griffith DA, Novak RJ. Decomposing malaria mosquito aquatic habitat data into spatial autocorrelation eigenvectors in a SAS/GIS® module. Trans GIS 2008; 12: 341-64.

[61] Jacob B, Muturi E, Funes J, *et al.* Using imaging technologies to control malaria. Imag Notes 2007; 3: 11-14.

[62] Jacob BG, Griffith DA, Gunter JT, *et al.* Spatial filtering specification for an auto-negative binomial model of *Anopheles arabiensis* aquatic habitats. Trans GIS 2008; 12: 243-59.

[63] Avery TE, Burkhart HE. Forest measurements. 4th ed. New York: McGraw Hill Book Co. 1994.

[64] Cressie NAC. Statistics for spatial data Revised ed. New York: Wiley 1993.

[65] Li X, Strahler AH. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. IEEE Trans Geosci Remote Sens 1992; 30(2): 276-92.

[66] Jensen JR. Remote sensing of the environment: an earth resource perspective. London, UK: Prentice Hall International Limited 2002.

[67] Haight FA. Handbook of the Poisson Distribution. New York: Wiley 1967.

[68] Neter J, Wasserman W, Kutner W. Applied linear statistical models. Richard D, Ed. Boston: Irwin 1990.

[69] Mwangangi J, Muturi EJ, Shililu J, *et al.* Environmental covariates of *Anopheles arabiensis* in a rice agroecosystem. J Amer Mosq Ass 2007; 23(4): 13-22.

[70] Chandler JA, Highton RB. The succession of mosquito species (Diptera: Culicidae) in rice fields in the Kisumu area of Kenya, and their possible control. Bull Entomol Res 1975; 6: 5295-302.

[71] Muturi EJ, Mwangangi J, Shililu J, *et al*. Mosquito species succession and physicochemical factors affecting their abundance in rice fields in Mwea, Kenya. J Med Entomol 2007; 44(2): 336-44.

[72] Jacob BG, Muturi EJ, Funes J, *et al.* Association between land cover and habitat productivity of malaria vectors in central Kenyan ricelands. *Acta Ento of China* 2007.

[73] Kaufman YJ, Tanre D, Remer LA, *et al.* Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imagind spectroradiometer. J Geo Phys Res 1997; 102(D14): 17051-67.

[74] Wanner W, Li X, Strahler AH. On the derivation of kernels for kernel-driven models of bidirectional reflectance. J Geo Phys Res 1995; 100(D10): 21077-89.

[75] McNairn H, Deguise JC, Pacheco A, *et al.* Estimation of crop cover and chlorophyll from hyperspectral remote sensing. 23rd Canadian Rem Sens Symposium; Sainte-Foy, Quebec, Canada2001.

[76] Peddle DR, Hall FG, LeDrew EF. Spectral mixture analysis and geometric-optical reflectance modeling of boreal forest biophysical structure. Remote Sens Enviorn 1999; 67(3): 288-97.

Received: February 27, 2011 Revised: April 25, 2011 Accepted: May 05, 2011