Methodology

Linking field-based ecological data with remotely sensed data using a geographic information system in two malaria endemic urban areas of Kenya

Thomas P Eisele*[1], Joseph Keating[1], Chris Swalm[2], Charles M Mbogo[3], Andrew K Githeko[4], James L Regens[5], John I Githure[6], Linda Andrews[7] and John C Beier[8]

Address: [1]Department of International Health and Development, Tulane University, School of Public Health and Tropical Medicine, New Orleans, LA, USA, [2]Department of Pharmacology, Tulane University, School of Public Health and Tropical Medicine, New Orleans, LA, USA, [3]Centre for Geographic Medicine Research, Kenya Medical Research Institute, Kilifi, Kenya, [4]Centre for Vector Biology and Control Research, Kenya Medical Research Institute, Kisumu, Kenya, [5]Institute for Science and Public Policy, Sarkeys Energy Center, University of Oklahoma, Norman, OK, USA, [6]Human Health Division, International Centre of Insect Physiology and Ecology, Nairobi, Kenya, [7]USRA, NASA Ames Research Center, Moffett Field, CA, USA and [8]Global Public Health Program, University of Miami, Miami, FL, USA

Email: Thomas P Eisele* - teisele@tulane.edu; Joseph Keating - jkeating@tulane.edu; Chris Swalm - cswalm@tulane.edu; Charles M Mbogo - cmbogo@kilifi.mimcom.net; Andrew K Githeko - AGitheko@kisian.mimcom.net; James L Regens - regens@ou.edu; John I Githure - jgithure@icipe.org; Linda Andrews - landrews@mail.arc.nasa.gov; John C Beier - jbeier@med.miami.edu

* Corresponding author

Abstract

Background: Remote sensing technology provides detailed spectral and thermal images of the earth's surface from which surrogate ecological indicators of complex processes can be measured.

Methods: Remote sensing data were overlaid onto georeferenced entomological and human ecological data randomly sampled during April and May 2001 in the cities of Kisumu (population $\approx$ 320,000) and Malindi (population $\approx$ 81,000), Kenya. Grid cells of 270 meters × 270 meters were used to generate spatial sampling units for each city for the collection of entomological and human ecological field-based data. Multispectral Thermal Imager (MTI) satellite data in the visible spectrum at five meter resolution were acquired for Kisumu and Malindi during February and March 2001, respectively. The MTI data were fit and aggregated to the 270 meter × 270 meter grid cells used in field-based sampling using a geographic information system. The normalized difference vegetation index (NDVI) was calculated and scaled from MTI data for selected grid cells. Regression analysis was used to assess associations between NDVI values and entomological and human ecological variables at the grid cell level.

Results: Multivariate linear regression showed that as household density increased, mean grid cell NDVI decreased (global F-test = 9.81, df 3,72, *P-value* = <0.01; adjusted R^2 = 0.26). Given household density, the number of potential anopheline larval habitats per grid cell also increased with increasing values of mean grid cell NDVI (global F-test = 14.29, df 3,36, *P-value* = <0.01; adjusted R^2 = 0.51).

Conclusions: NDVI values obtained from MTI data were successfully overlaid onto georeferenced entomological and human ecological data spatially sampled at a scale of 270 meters × 270 meters.

Results demonstrate that NDVI at such a scale was sufficient to describe variations in entomological and human ecological parameters across both cities.

Background

At present, roughly half of the world's population lives in urban environments. Furthermore, it is expected that virtually all the population increases of underdeveloped regions of the world will be concentrated within small and medium sized urban areas over the next thirty years [1]. This high rate of urbanization will undoubtedly have profound implications for the epidemiology of malaria, as well as other vector-borne diseases in sub-Saharan Africa [2-7].

High resolution passive remote sensing systems, such as the Multispectral Thermal Imager (MTI) satellite, provide detailed spectral and thermal images of the earth's surface from which surrogate ecological indicators of complex processes can be measured [8]. Such remote sensing systems may, therefore, prove to be useful in assessing ecological changes within vast, highly heterogeneous urban areas. In many developing countries, the scarcity of resources, coupled with the need for regular surveillance, requires the development of novel methods for the collection and analysis of ecological data.

Since the mid 1980's, there has been a strong interest in the application of passive remote sensing from satellites to further our understanding of the epidemiology of vector-borne diseases, especially malaria [9-16]. The normalized difference vegetation index (NDVI), which expresses the abundance of actively photosynthesizing vegetation, or "greenness" [8], has been of particular interest in mapping both spatial and temporal relationships between the environment and malaria incidence [11,12,14]. However, remote sensing data obtained from satellites up to this point have primarily been limited to assessing associations of environmental conditions, such as NDVI, with attributes of vector-borne diseases within non-urban areas, and often over spatial scales unsuitable for measuring ecological conditions within heterogeneous and complex environments such as cities. The MTI satellite provides visible and near infrared data at a spatial resolution of five meters, and near, medium and thermal infrared data at a spatial resolution of 20 meters, which should prove sufficient for assessing ecological variations within urban environments. Furthermore, MTI data at such spatial scales can be easily coupled with georeferenced field-based data, thus enabling it to be used as a proxy for community level conditions that may be too expensive or logistically difficult to collect in areas with limited surveillance or research resources.

A method of overlaying NDVI values, derived from MTI data, on to georeferenced entomological and human household ecological data randomly sampled at a spatial scale of 270 meters × 270 meters within the cities of Kisumu and Malindi, Kenya are described within this paper. Additionally, an attempt is made to demonstrate and validate the utility of this methodology by testing the hypothesis that NDVI derived from MTI data is sufficient to describe variation in entomological and human ecological parameters at such a scale within these urban environments.

Methods

Study sites

The study sites of Kisumu and Malindi in Kenya have been described in detail elsewhere [7,17]. Briefly, Kisumu is Kenya's third largest city with a population of approximately 320,000. The city is located on Lake Victoria within Nyanza Province, 10 km south of the equator, approximately 1,100 meters above sea level (Figure 1). Mean daily temperatures range from 18° to 30°C, while the average annual rainfall typically varies between 1,100 and 1,300 millimeters. Kisumu town is comprised of residential, commercial, and industrial areas, with undeveloped land and vegetation both within and around the urban sub-locations. Most roads within Kisumu are paved with covered engineered drainage systems lining both sides of the streets, although some roads and paths, both within and around the city, consist of dirt. Seventy-five percent of the urban population has access to piped water [18], although water vendors are also common. Malaria transmission within outlying rural areas of Kisumu occurs throughout the year at intense levels, with *Anopheles gambiae s.l.* and *Anopheles funestus* the primary malaria vectors in this area of western Kenya [19-21].

Malindi is Kenya's tenth largest city with a population of approximately 80,500. The city is located on the shore of the Indian Ocean in Coast Province (Figure 1). Mean daily temperatures range from 22° to 30°C, while average annual rainfall varies between 75 and 1,200 millimeters along the Kenyan coast. Malindi town is made up of residential, commercial, and agricultural areas, with tourism related activities dominating the areas closest to the coast. The center of town has functional engineered drainage systems lining both sides of paved roads. Roads along the coast are also paved. Many roads and paths within the town are a mixture of sand and dirt. Sixty percent of the urban population has access to piped water [18], although shallow garden wells are also used to obtain household
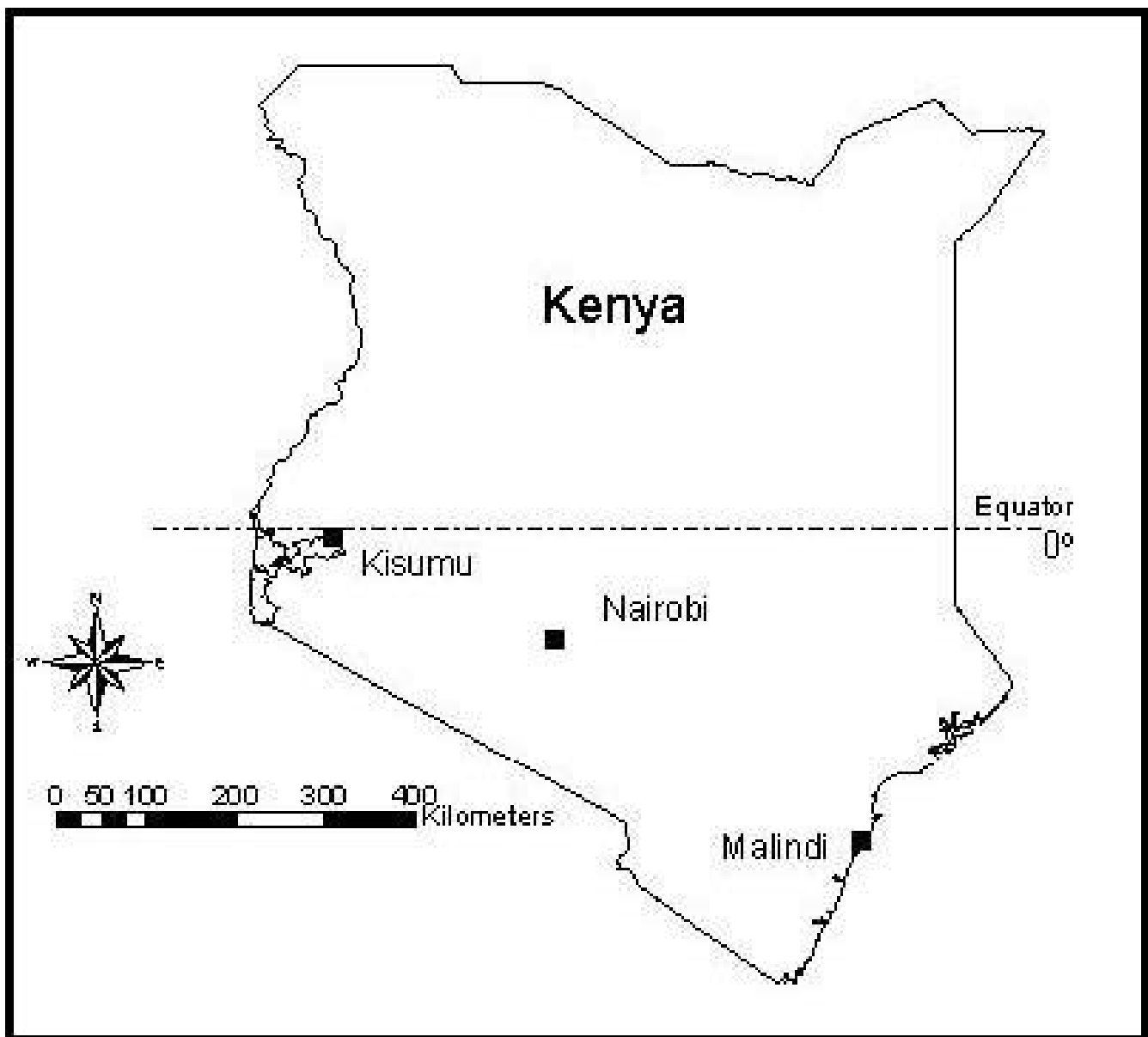
Figure 1
Map of Kenya showing Kisumu and Malindi *

water. Malaria transmission along the coast of Kenya has been characterized as low and seasonal, with *An. gambiae s.l.* and *An. funestus* the primary malaria vectors [22].

Sample frame development
The sampling strategy used for the collection of larval site and household data was developed for an earlier research project and has been described in detail elsewhere [7]. Briefly, the goal was to generate a sample frame to support data collection efforts and subsequent hypothesis testing

across multiple disciplines with minimal house and larval site selection bias. Base maps of major roads and hydrography were created using ArcView 3.2® (Environmental Systems Research Institute, Redlands, CA), a geographic information system (GIS), on which a series of 270 meter × 270 meter grid cells were overlaid to generate spatial sampling units for multi-disciplinary data collection efforts. Grid cells falling within the urban context were located in the field, described, and stratified based on the level of planning and drainage present per grid cell,
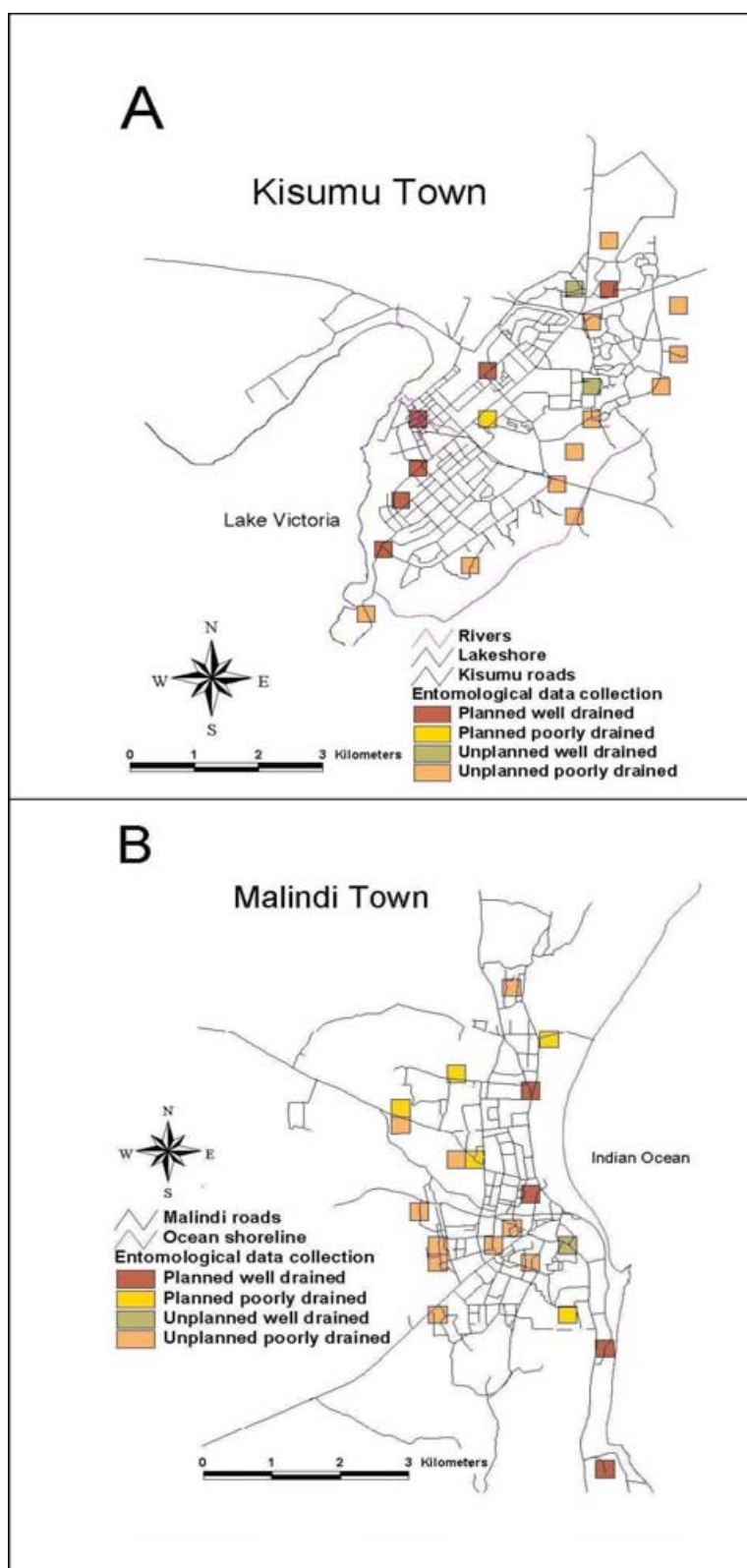
Figure 2
Randomly selected grid cells in Kisumu (A) and Malindi (B), Kenya, for entomological data collection, by four strata of planning and drainage typology *
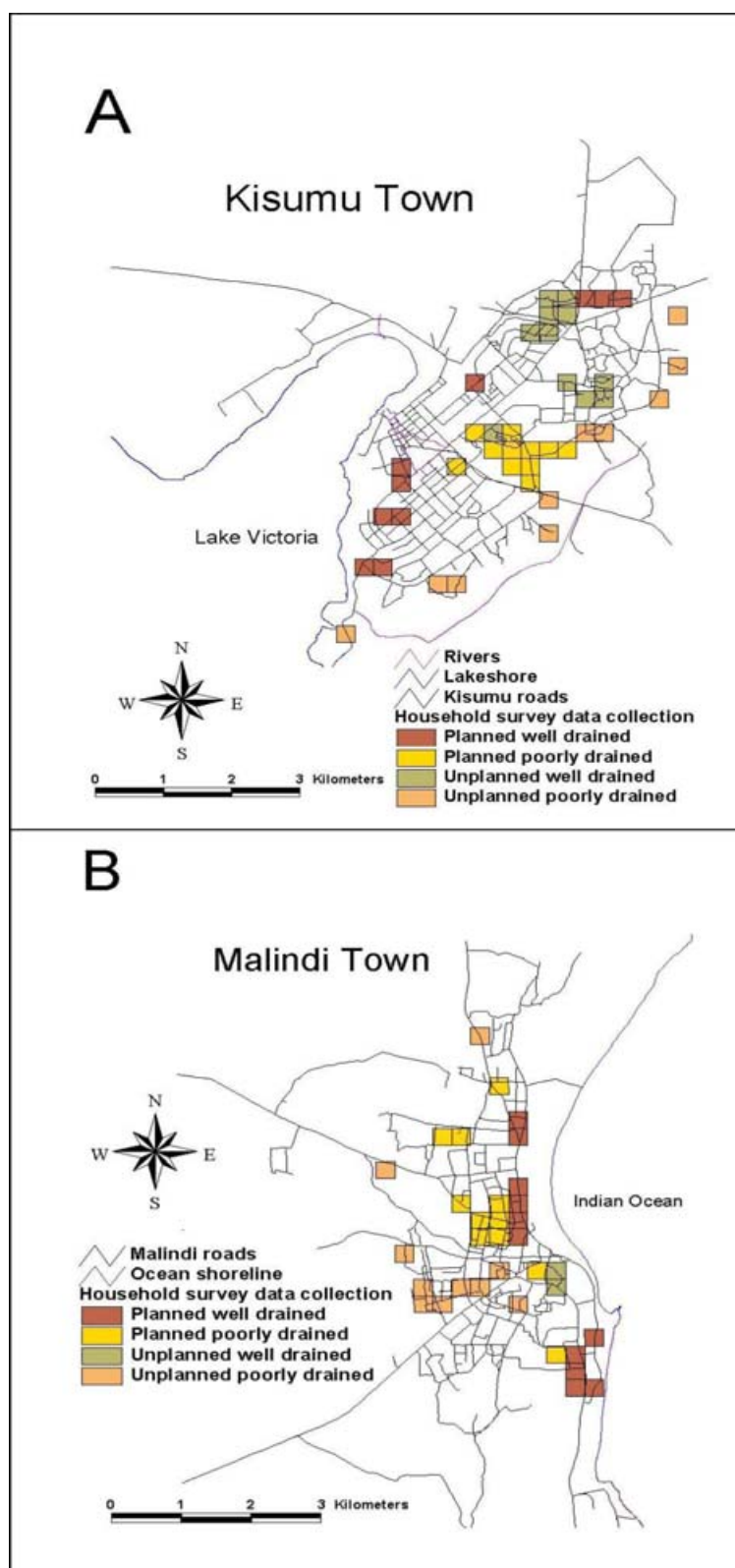
Figure 3
Randomly selected grid cells in Kisumu (A) and Malindi (B), Kenya, for household survey data collection, by four strata of planning and drainage typology *

defined as: 1) planned, well drained; 2) planned, poorly drained; 3) unplanned, well drained; and 4) unplanned, poorly drained. Although a rural stratum in Kisumu and a peri-urban stratum in Malindi were identified as part of the original sampling frame, they have been excluded here because they could not be successfully linked with the MTI data through unique identifiers. Probability proportional to size sampling, based on the proportion of grid cells within each stratum, was used to randomly select 20 grid cells in Kisumu and 20 grid cells in Malindi for entomological sampling. These selected grid cells also served as a guide for household level data collection efforts described below. The number of grid cells selected for data collection was a function of time and operational constraints. As well, the size of the grid cell (i.e. 270 meters × 270 meters) was established to coincide with a 9 pixel by 9 pixel (30 meter resolution) LANDSAT Thematic Mapper (TM) image for an earlier study, and is not the subject of this demonstration. Figures 2 and 3 illustrate the grid cells used in this analysis for entomological and household survey data collection, respectively, by four strata of planning and drainage typology within Kisumu and Malindi.

Entomological data collection

All accessible water bodies within the grid cells selected within Kisumu (20 grid cells) and Malindi (20 grid cells) were located, georeferenced, described and sampled for mosquito larvae during April and May 2001 (Figure 2), as described in detail elsewhere [7]. Briefly, all water bodies were visited to assess the potential for anopheline ovipositioning and subsequent larval development. All such identified water bodies were georeferenced with a Trimble (Sunnyvale, CA) global positioning system (GPS) receiver. Water bodies were visually inspected for the presence of larvae. When present, standard sampling techniques were used to collect, preserve and transport specimens to the laboratory for identification [23]. Water bodies were also described in terms of surrounding environmental and habitat characteristics.

Aquatic habitats were considered potential larval sites if at least one anopheline mosquito larva was present, or if no anopheline larvae were present but the water body was within 20 meters of an active anopheline larval site and similar with respect to key characteristics of substrate type, size and water quality and depth. Although the inclusion criterion for potential larval sites was general, it is recognize that a plethora of factors interact on multiple scales to effect ovipositioning preferences and subsequent larval development, and that not all potential larval sites will produce adult mosquitoes.

Household data collection

In both cities, additional grid cells were randomly selected for the administration of a household questionnaire during April and May 2001, as described in detail elsewhere [7,17]. The purpose of the survey was to collect and compare data relative to socioeconomic status, mosquito avoidance behaviours, and knowledge of mosquito life cycles, between the respective strata of planning and drainage. Thus, the number of households selected per strata was not proportional to the actual number of households, nor grid cells, per strata. In many instances, households or owners were either absent or unwilling to be interviewed within specific grid cells. This necessitated the random selection of adjacent grid cells of the same stratification typology to reach the target sample size of 100 households per stratum. Households sampled were randomly selected using the center of the grid cell as a reference point. In Kisumu, 411 households from 42 grid cells were chosen within the four respective strata of planning and drainage. In Malindi, 380 households from 34 grid cells were chosen within the four respective strata of planning and drainage (Figure 3). Refer to Macintyre et al (2003) for further description of the household-level socioeconomic data.

In addition to entomological and household data collection, the total numbers of occupied households contained within all selected 270 meter × 270 meter grid cells were obtained. This measure, defined as household density per grid cell, was intended to serve as a community level surrogate for population density per grid cell. This count also served as a denominator for the calculation and application of sampling weights within each grid cell. Two independent observers performed household counts. When counts yielded different results for the same grid cell, two additional counts were performed and the average of the four counts was used.

The institutional review boards of the Kenya Medical Research Institute (Nairobi, Kenya) and Tulane University (New Orleans, LA) approved the study protocol and questionnaire for the household data collection.

Multispectral Thermal Imager data collection

The MTI satellite was used to collect the remotely sensed data. The MTI satellite programme is sponsored by the U.S. Departments of Energy and Defense, and coordinated by Sandia National Laboratories, Los Alamos National Laboratory, and the Savannah River Technologies Center. While the primary objective of the MTI programme involves national defense, a secondary objective is to collect ground data to validate the algorithms used in the calibration of the sensors, and to verify the accuracy and interpretability of the remotely measured data for research applications.

The MTI satellite is designed to collect, compress, and store six, 12 × 12 km images per day in 15 spectral bands

ranging from 0.45 to 10.70 µm (visible blue through long wave infrared) [24]. The four shortest wavelength bands (bands A-D) are at five meter spatial resolution, with the remaining bands are at a spatial resolution of 20 meters. Bands A-D range from visible blue through very near infrared, while the remaining bands (E-N) range from near infrared to long-wave infrared and thermal.

MTI data were acquired for Kisumu and Malindi Districts for the dates of 20 February, 2001 and 4 March, 2001, respectively, after obtaining permission to join the MTI Users Group. MTI data were collected between 0900 and 1000 coordinated universal time at both sites. The target coordinates were -0.282002 degrees latitude and 34.7700 degrees longitude for Kisumu and -3.31300 degrees latitude and 40.13200 degrees longitude for Malindi. Figures 4 and 5 show infrared color composite images derived from the MTI data for the areas of Kisumu and Malindi, respectively. This analysis used bands C and D, collected at nadir only, to calculate NDVI for association with selected grid cells to investigate environmental differentiations over areas of human-ecological heterogeneity. The results of additional analyses using thermal bands, as well as additional grid cells, will be presented in a series of subsequent papers.

Geoprocessing, calculation of NDVI, and overlaying onto field-based data

MTI data subsets, comprising two scenes, one for Kisumu and one for Malindi, were processed to a level1b_r_coreg product in hierarchical data format (HDF) prior to being obtained from Sandia National Laboratories. This level of processing maps the raw satellite imagery to a grid based on the WGS84 ellipsoid. The HDF data sets were converted to ERDAS *Imagine* format files for use with ArcView®. ArcView 3.3® was used to view the data subsets. Bands A-D were visually inspected for clarity, variability and ultimate usefulness for this demonstration. The MTI data were then georeferenced to the Universal Transverse Mercator (UTM) coordinate system and cross-referenced with five control points from existing LANDSAT TM data, which were already projected to UTM. The MTI data were georeferenced to five meter spatial resolution using a 'nearest neighbor' resampling setting. Once GPS based data became available from the fieldwork in Kisumu and Malindi, a projection was performed, which converted the GPS based data from the geographic latitude and longitude coordinate system to the UTM coordinate system, to correspond with the MTI data subsets. False easting and false northing settings were applied to improve the registration of the respective data sets.

The *Image Analysis* extension of ArcView 3.3® was used to perform NDVI calculations of the ERDAS *Imagine* formatted files. NDVI was calculated as (Band D - B and C) /

(Band D + Band C). The MTI bands D and C correspond to the infrared and red wavelengths, respectively. The MTI band wavelengths ranged from 0.62–0.68 µm for the red and 0.76–0.86 µm for the infrared. The NDVI calculation results in an ERDAS *Imagine* floating-point format file, with NDVI values ranging from -1 to 1. In order to overlay these data on the existing base maps and selected grid cells, the MTI data were added to the ArcView® project file for further processing and viewing. The cartographic information for the two base maps was stored as separate shape files within the ArcView® project file, and overlaid with the MTI data as a layer for each city.

The different modules within the ArcView® software can use and display the ERDAS *Imagine* format files in different ways. As well, the *Imagine* format files are not always accessible to all modules of the ArcView® software and its extensions. Thus the decision was made to convert the *Imagine* format files to Arc/Info GRID files and recalculate the NDVI values using the following calculation in ArcView *Spatial Analyst*: NDVI = ((([Band D] - [Band C]) / ([Band D] + [Band C]).Float). This produced an Arc/Info GRID format file with a floating-point data range of -1 to 1. Randomly selected locations from both file formats were compared to ensure correct calculation of NDVI in *Spatial Analyst*. The validation was performed by identifying and recording X,Y coordinates from the *Imagine* format data images, recording the NDVI values at these locations, and then pointing to the corresponding locations in the Arc/Info GRID format file and comparing the values. All such readings were identical. This process is useful for calculation validation as well as for verifying floating-point values. A cross-tabulation was performed in *Spatial Analyst*, using the "summarize by zone" function, to evaluate the randomly selected grid cells with entomological and household data.

As a result of this process, a database was created for each city with the mean, minimum, maximum, and standard deviations for NDVI data aggregated to the 270 meter × 270 meter grid cell level. The mean NDVI value of all points within each selected 270 meter × 270 meter grid cell was used in this analysis. Adding all NDVI pixel values within the respective grid cell, and dividing that number by the total number of pixels falling within the grid cell, was the process used to calculate the mean NDVI value per grid cell. The NDVI datasets for Kisumu and Malindi were then merged with the entomological and household datasets using unique identifiers for each selected grid cell. To aid with interpretation, NDVI values were then scaled to a range of 0 to 200 as follows: scale NDVI = 100(NDVI + 1). Although it is recognized that in aggregating the MTI data up to the 270 meter × 270 meter grid cell level important information may have been lost, this was necessary because entomological and household level data were

Figure 4
Infrared color composite derived from the MTI data for the area of Kisumu, Kenya *

Figure 5
Infrared color composite derived from the MTI data for the area of Malindi, Kenya *
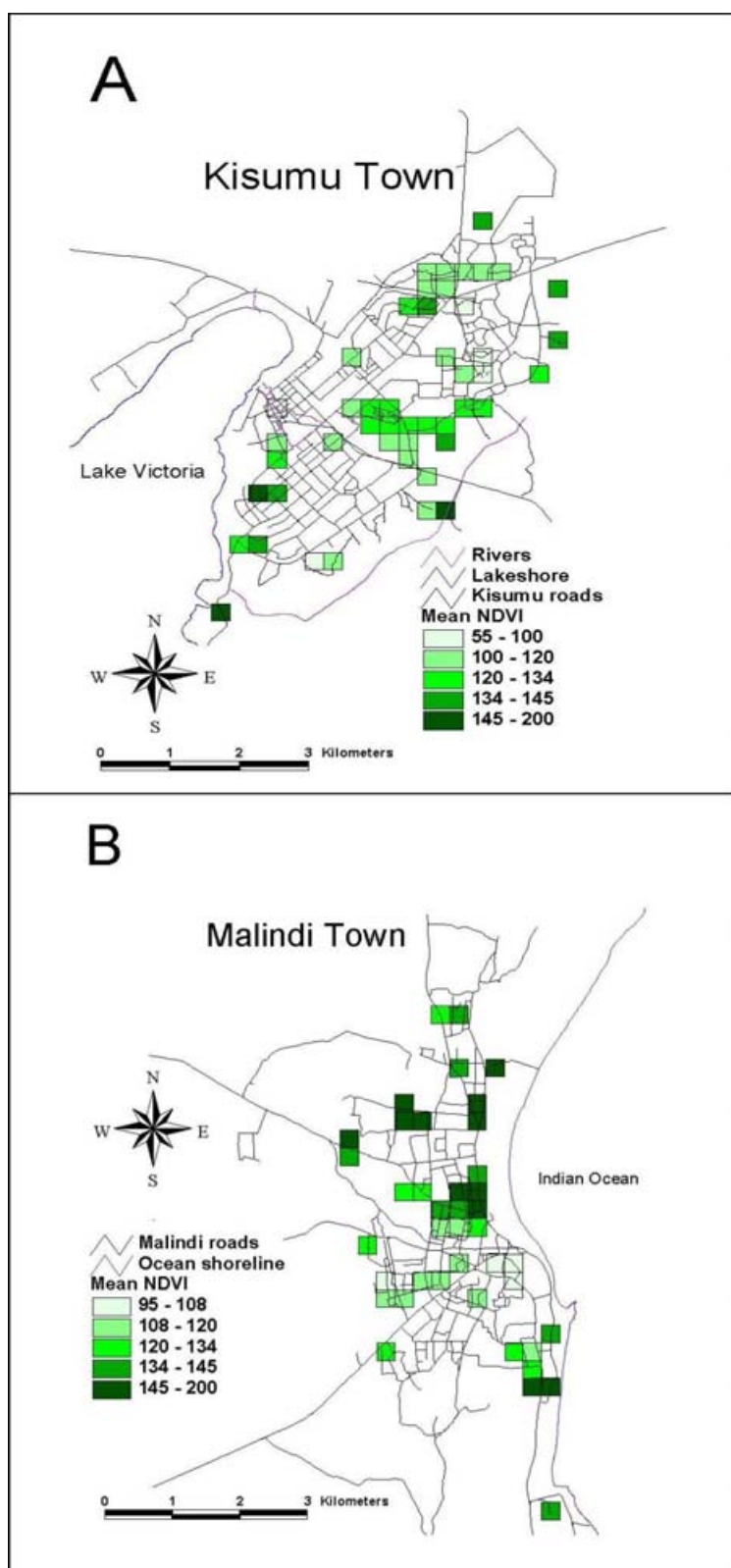
Figure 6
Mean values of NDVI for selected grid cells in Kisumu (A) and Malindi (B), Kenya, by four strata of planning and drainage typology *

collected at this scale. Figure 6 illustrates the distribution of the resultant mean scaled NDVI values for each randomly selected grid cell within Kisumu and Malindi. The values are categorized based on the overall range observed for each city, and both entomological and household surveyed grid cells are included.

Images were taken with a digital camera at points within Kisumu and Malindi during field-based data collection. Latitude and longitude were also recorded at these points using GPS. Six of these images, coupled with their respective GPS coordinates, were used to assist with the interpretation of NDVI values at five meter resolution within selected grid cells.

Data management and statistical analysis

ScienceOrganizer, a common internet-based repository designed by the U.S. National Aeronautic and Space Administration to enhance information storage, organization, and access between research team members at different locations, was used to upload, store and access the field-based data. SAS 8.01 (SAS Inst, North Carolina) was used for all data cleaning, management and statistical analyses.

To validate the method of overlaying MTI data onto georeferenced field-based data described within this paper, relationships between mean grid cell NDVI and field-based data relating to entomological and human activity were investigated. Data analysis was performed on a finite number of entomological and household variables that were expected to correlate in a specific direction with NDVI in order to demonstrate the validity of the method described within this paper. Please note that it is not the intension of this paper to assess the relationships between NDVI and field-based data for the purpose of developing predictive or explanatory models in themselves, but rather to validate that NDVI derived from MTI data are adequate to describe urban environment such as these at the grid cell level. A two-sided *P-value* of <0.05 was considered statistically significant.

To ensure a probability sampling design with respect to the number of grid cells per strata of planning and drainage, sampling weights were assigned to all grid cell level data (NDVI, number of potential anopheline larval sites and household density per grid cell). Sampling weights were calculated for each grid cell based on its probability of selection as follows: grid cell weight = 1 / (number grid cells selected within strata *x* / total number grid cells in strata *x*). Such weights were used in generating all descriptive statistics and subsequent statistical analyses at the grid cell level.

As stated, the selection of households was not proportional to the number of grid cells per strata of planning and drainage, nor the number of households per grid cell. To restore a probability sampling design, sampling weights were assigned to the household level data as follows: household level weight = 1 / ([number grid cells selected within strata *x* / total number grid cells in strata *x*] * [number of households selected within grid *y* / number of households within grid *y*]). Such weights were used in all subsequent statistical analyses at the household level.

Spearman's correlation coefficients were calculated to test the association between mean grid cell NDVI and the four respective strata of planning and drainage, within Kisumu and Malindi. Using weighted data, Pearson's correlation coefficients were also calculated to test the associations between mean grid cell NDVI and the number of potential anopheline larval habitats per grid cell, and household density per grid cell, within Kisumu and Malindi.

Multivariate linear regression was used to further investigate the potential association between mean grid cell NDVI and household density and the number of potential anopheline larval habitats per grid cell, while controlling for potential confounders. Such grid cell level data for Kisumu and Malindi were combined in order to investigate if such relationships were homogeneous across study sites, as well as to increase statistical power. As it was hypothesized that NDVI should increase with decreasing household density, a regression model was established with mean grid cell NDVI as the dependent variable and household density as the predictor variable of interest. The potential confounding effects of study site and planned versus unplanned drainage strata were controlled for in the model. As it was further hypothesized that the abundance of potential anopheline larval habitats should increase with increasing NDVI, a regression model was established with the number of potential anopheline larval habitats per grid cell as the dependent variable and mean grid cell NDVI as the predictor variable of interest. As it has been shown previously that grid cell household density is significantly associated with the number of potential anopheline larval habitats per grid cell within these two cities [7], the potential confounding effect of household density was controlled for in this model.

Multivariate logistic regression was also performed to investigate the relationship between mean grid cell NDVI (dichotomized as high or low) and the presence or absence of cultivated agricultural fields surrounding a house. As it was hypothesized that the odds of there being cultivated agricultural fields about a house should increase if mean grid cell NDVI was high, a logistic regression model was established with the presence/absence of cultivated agricultural fields surrounding a house as the

dependent variable and grid cell NDVI as the predictor variable of interest. The potential confounding effects of planned versus unplanned drainage strata, study site and low versus high household wealth were adjusted for in this model. As there was no significant interaction between study site and grid cell NDVI, data for Kisumu and Malindi were combined for this analysis. As this model included data at both the household and community levels, standard errors were empirically estimated using GEE (generalized estimation equation) methods.

Results
Results from the validation exercise suggest that the MTI data at a spatial resolution of five meters can accurately describe the level of "greenness" in the area. Figure 7 illustrates the digital images and corresponding NDVI values for three sites in Kisumu and three sites in Malindi. The high NDVI value detected for images A and C are expected, as both sites are in residential areas with high levels of vegetation. However, the NDVI value detected for image F is unexpected. The site was a swimming pool, constructed with concrete and tile. It is possible that the satellite imager was measuring vegetation in the surrounding area, as this site was found along the shoreline area with an abundance of ornamental vegetation both within and around neighboring hotel compounds. As well, the level of error associated with the registration process and Trimble GPS collected data points may be such that the true site is 0 to 15 meters from the swimming pool. Sensor saturation could also be responsible for this anomalously high value. Images B and E are both void of vegetation with high levels of metal and concrete, respectively, hence the low values. Site D illustrates the value when water is present with little or no vegetation around the site. In general, these values, with the exception of the swimming pool site, are within the expected range of NDVI values for the respective site characteristics.

Values for NDVI obtained from the MTI satellite were successfully aggregated and overlaid onto georeferenced field-based data for all selected 270 meter × 270 meter grid cells within Kisumu and Malindi. Mean grid cell NDVI values ranged from a low of 80.4 to a high of 168.3 across all sampled grid cells in both study sites (Mean NDVI = 122.7, standard deviation = 18.1). The mean NDVI was 121.3 within Kisumu and 128.0 within Malindi across sampled grid cells (Table 1). In both cities, areas that were unplanned but well drained had the lowest mean NDVI values. While unplanned poorly drained areas had the highest mean NDVI value in Kisumu (123.1), planned well drained areas had the highest mean NDVI value in Malindi (141.9), among sampled grid cells. There were no significant correlations between mean NDVI and the respective strata within Kisumu (n = 4,

Spearman r = 0.20, *P-value* = 0.80) or Malindi (n = 4, Spearman r = -0.80, *P-value* = 0.20).

A univariate analysis showed mean grid cell NDVI and household density to be significantly correlated in a negative direction within both Kisumu (Pearson r = -0.555, *P-value* = <0.01) and Malindi (Pearson r = -0.453, *P-value* = 0.01) (Table 2). Further analysis with a multivariate linear regression model confirmed this relationship, as household density was found to be a significant factor affecting mean grid cell NDVI in an inverse manner, after controlling for strata and study site (Table 3). This relationship did not differ by study sites, as there was no significant interaction between study site and household density (Partial F = 0.127 df 1,71, *P-value* > 0.10). This model was significant (global F-test = 9.81, df 3,72, *P-value* = <0.01) and explained approximately 26% of the variation in mean NDVI values across sampled grid cells.

Without controlling for confounders, there was no significant correlation between mean grid cell NDVI and the number of potential anopheline larval sites per grid cell within Kisumu (Pearson r = 0.267, *P-value* = 0.26) or Malindi (Pearson r = -0.319, *P-value* = 0.17) (Table 2). However, subsequent analysis with multivariate linear regression showed mean grid cell NDVI to be a significant factor affecting the abundance of potential anopheline larval habitats across grid cells in a positive direction, after controlling for household density (Table 4). This relationship was consistent across both study sites (Partial F-test with addition of study site*NDVI: F = 0.133 df 1,34, *P-value* > 0.10). This model was highly significant (global F-test = 14.29, df 3,36, *P-value* = <0.01), explaining over half (51%) of the variation in the number of potential anopheline larval habitats across sampled grid cells. Mean grid cell NDVI was responsible for 9% (0.044) of the total adjusted R^2 of 0.51.

Multivariate logistic regression showed that the odds of there being cultivated agricultural fields around a house increased significantly for those located within grid cells with high NDVI, as compared to grid cells with low NDVI, after adjusting for planning and drainage, study site and household wealth (n = 76, OR = 2.0, *P-value* = 0.05).

Discussion
NDVI values obtained from MTI data were successfully overlaid onto georeferenced field-based entomological and human household ecological data at a spatial scale of 270 meters × 270 meters within the cities of Kisumu and Malindi, Kenya. These results demonstrate that mean grid cell NDVI at such a scale was sufficient to describe variations in specific entomological and human ecological parameters across both urban environments. Such relationships were statistically significant, and most
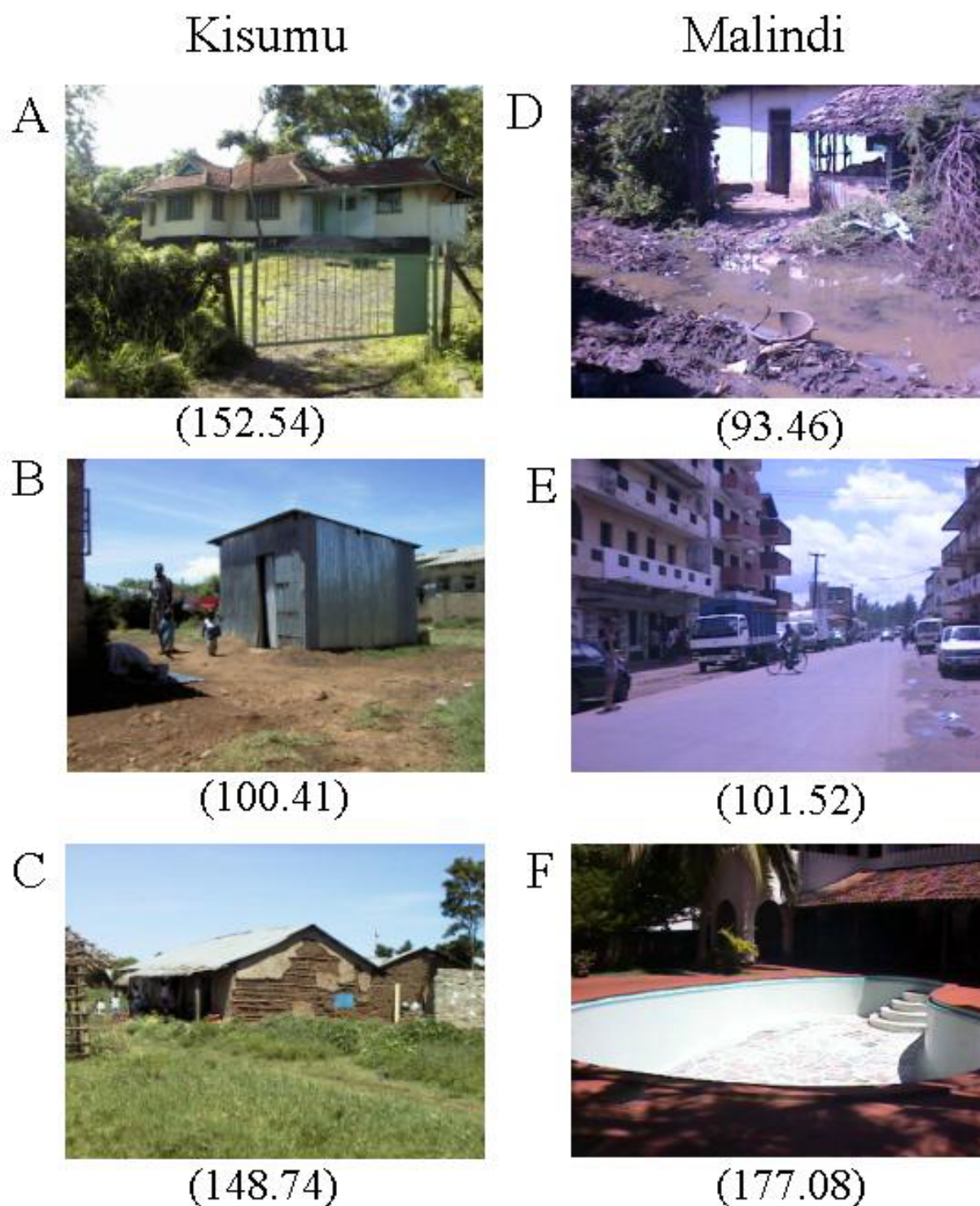
Figure 7
Digital images illustrating NDVI values for specific sites in Kisumu and Malindi, Kenya. *

Table 1: Variance in mean NDVI by planning and drainage typology within Kisumu and Malindi, Kenya.

	Grids (n)	Mean NDVI†	Standard deviation
Drainage typology			
		Kisumu*	
Planned, well drained	10	121.3	21.48
Planned, poorly drained	12	117.2	4.01
Unplanned, well drained	10	111.0	11.01
Unplanned, poorly drained	10	123.1	28.96
Total	**42**	**121.3**	**18.12**
		Malindi*	
Planned, well drained	11	141.9	10.67
Planned, poorly drained	10	137.1	16.70
Unplanned, well drained	2	105.5	1.47
Unplanned, poorly drained	11	116.9	13.05
Total	**34**	**128.0**	**17.19**

†Estimated means based on weighted data at grid cell level. *No significant correlation between NDVI and drainage typology for either Kisumu or Malindi (Kisumu: n = 4, Spearman r = 0.20, *P-value* = 0.80; Malindi: n = 4, Spearman r = -0.80, P-value = 0.20)

Table 2: Correlation of mean NDVI against number of potential anopheline larval sites, and household density, per grid cell: Kisumu and Malindi, Kenya

	Grids (n)	Correlation coefficient (r)†	*P-value*
Variable			
		Kisumu	
Household density*	42	-0.555	<0.001
Number of potential anopheline larval sites**	20	0.267	0.255
		Malindi	
Household density*	34	-0.453	0.007
Number of potential anopheline larval sites**	20	-0.319	0.171

†Pearson's correlation coefficient used with weighted data at grid cell level. *Collected during household sampling. **Collected during larval data collection.

Table 3: Summary statistics of linear regression model relating household density to mean grid cell NDVI: Kisumu and Malindi, Kenya

	Coefficient	Standard error	*P-value*
Model predicting NDVI (Y)			
$(Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e)$			
β_0: Constant	138.947	4.219	<0.001
X_1: Household density	-0.073	0.017	<0.001
X_2: Planned/unplanned-drainage	-1.217	4.126	0.769
X_3: Kisumu/Malindi	-1.122	3.910	0.775
Adjusted R^2: 0.26			
Global F test (P-value): 9.81 (<0.001)			
n = 76*			

*Collected during household survey, Kisumu and Malindi combined.

Table 4: Summary statistics for linear regression model relating mean grid cell NDVI to potential anopheline larval habitat abundance: Kisumu and Malindi, Kenya

	Coefficient	Standard error	*P-value*
Model predicting potential anopheline larval habitat abundance (Z) $(Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2{}^2 + e)$			
β_0: Constant	-3.256	2.150	0.129
X_1: NDVI	0.031	0.015	0.047
X_2: Household density	0.055	0.012	<0.001
$X_2{}^2$: Household density squared	<-0.001	<0.001	<0.001
Adjusted R^2: 0.51			
Global F test (P-value): 14.29 (<0.001)			
n = 40*			

*Collected during larval sampling, Kisumu and Malindi combined.

importantly were in the hypothesized directions. Moreover, such relationships between mean grid cell NDVI and field-based entomological and human ecological data were consistent across both Kisumu and Malindi, which are actually quite different cities with regards to ecology, culture, history and economics. However, due to small sample sizes, possible mis-registration of the MTI data, and information lost as a result of aggregating data, results from the analyses used to test relationships between grid cell NDVI and field-based data should be interpreted with caution.

As hypothesized, it was found that as household density increased, mean grid cell NDVI – or the amount of greenness in the area – decreased, after adjusting for study site and planning and drainage. It was also found that given household density, the number of potential anopheline larval habitats increased with increasing values of mean grid cell NDVI. This suggests that both the abundance of potential larval sites and the amount of greenness increased with the propensity of areas within a grid cell to hold soil moisture, which is due to a multitude of factors. And lastly, as hypothesized, it was found that the odds of there being cultivated agricultural fields around a house doubled if it was located within a grid cell with high, as opposed to low, mean NDVI. This suggests that it is possible to use NDVI as a surrogate for community level factors such as the presence of urban farming at a scale of 270 meters × 270 meters within an urban context. This is important as urban farming likely provides ample aquatic habitats for mosquitoes. Furthermore, it has been shown that peri-urban areas, which may include urban areas with substantial amounts of subsistence farming, have higher levels of annual malaria transmission compared to strictly urban areas [2]. Thus the use of surrogate ecological variables such as NDVI may allow for

assessment of urban areas in terms of their overall propensity to harbor mosquito populations. Because remote sensing data is often available at a bi-weekly interval, changes in the ecological status of an area can also be determined over time, which is important from a surveillance standpoint within urban environments.

Overall, mean grid cell NDVI did not correlate consistently with levels of planning and drainage. Grid cells defined as planned and well drained within both Kisumu and Malindi had relatively high mean NDVI values. This result was not unexpected as such areas are generally more affluent and have relatively lower household densities, thus there is often more planned vegetation such as planted trees, ornamental plots, and small gardens. Unplanned well drained areas had the lowest mean values for NDVI in both Kisumu and Malindi. This result was also not surprising as such areas were typically situated on slopes, or hillsides, which did not permit the accumulation of standing water.

The random selection of grid cells, proportional to the number of grid cells per strata of planning and drainage, for collecting data on mosquito larvae and potential anopheline larval habitats proved sufficient to generate a relatively unbiased sample of such elements. Such a sampling strategy is appropriate because pools of water, which make up mosquito larval habitats, can be assumed to be approximately randomly distributed across a given space. However, the spatial sampling technique used in this demonstration was not an efficient design for the selection of households, as it cannot be assumed that households are randomly distributed across space, especially with respect to strata of planning and drainage. It was therefore necessary to assign sampling weights to all households *post hoc* based on estimates of the number of

grid cells selected per strata (fist stage), and the number of households selected from within a chosen grid cell (second stage). However, a more appropriate method for assigning sampling weights to the household-level data would have been to base the first stage on the number of households per strata, although such data were not available at the time of this analysis. For this reason, the household sample merely approximates a probability sampling design across the planning and drainage strata, and over both study sites combined. It is suggested that households not be spatially sampled, but rather randomly sampled from an enumeration of households, when available. When such a sampling frame does not exist, multistage sampling techniques should be derived from standard methods [25] pertaining to estimates of the number of households per desired strata and the number of households per selected cluster to produce a relatively unbiased sample. Households selected in this way can then be georeferenced using GPS in order to be linked with remotely measured data.

The MTI data used in this research were obtained for a period of time approximately two to three months prior to the collection of field-based household data. Such temporal disparities may have confounded our observed relationships, although the exact nature of such confounding is unknown. Furthermore, it may be that the utility of NDVI to adequately describe urban conditions is limited to periods of time when there is enough precipitation to yield sufficient variations in greenness across relatively small areas.

An additional issue worth discussing is the nature of any spatial autocorrelation that may exist within the data. Although many continuous surfaces contain a certain degree of autocorrelation, and near points may be more alike in value than points at a further distance, no test for spatial or residual autocorrelation was performed. While the importance of establishing a level of independence within the data prior to performing linear regression analysis is recognized, the unit of analysis at the aggregated level and the small sample size precluded further investigations into the spatial structure of the data. However, the purpose of this paper was not to build an exhaustive predictive model that accounts for the effect of latitude and longitude on estimated coefficients and observed relationships; rather the purpose was to describe and demonstrate the utility of a method for integrating multidisciplinary data in a public health capacity.

It is recognized that interactions and feedbacks between humans and their environments must be accounted for to successfully describe, and thus understand, urban ecosystems [26]. These results demonstrate that NDVI at such a scale may prove sufficient to help describe urban

environments in terms of both natural and human-influenced ecological characteristics that may affect malaria transmission. According to these findings, urban areas with higher NDVI may likely have lower household/population densities and have increased potential to harbor anopheline larval habitats. In many instances within the context of this research, these areas were of high affluence, with patches of ornamental vegetation and small gardens both within and around households. On the other hand, urban areas with lower NDVI may likely represent areas with high household/population density, such as slums constructed mostly of wood and sheet metal, and urban business districts, such as parking lots and buildings constructed mostly of concrete. Such 'brown' areas would most likely not hold much potential to harbor natural habitats for anopheline mosquito larvae. However, it is unknown to what extent *Anopheles* mosquitoes are adapting to urban conditions [2]. In the presence of either genotypic or behavioural adaptation of the anopheline to urban environments, ovipositioning preferences and larval development strategies may change [3,27]. Thus what holds true in rural environments may not necessarily hold true in the context of the urban ecosystem, thus areas of low NDVI may also be important in terms of *Anopheles* habitat development.

Remote sensing data such as NDVI represent a vector of community level variables, and thus hold potential to supplement data collected from georeferenced household surveys within urban areas. Additionally, remote sensing data at similar scales could potentially provide pubic health officials with valuable information with regards to urban neighborhoods that may be at increased risk of malaria outbreaks and would likely benefit from integrated vector control. Although the effectiveness of larval control has not been demonstrated across a range of urban environments in Africa, its use as part of an integrated vector management system is warranted in the presence of heightened malaria parasite transmission. This demonstration provides some evidence that remote sensing data have the potential to be used as a valuable surveillance tool for both public health research groups and local organizations involved with vector-borne disease control within urban areas, although to what extent remotely sensed data can definitively identify areas with an increased propensity to harbor adult or larval stage anopheline mosquitoes remains unknown.

Authors' contributions

TE helped conceive this research, performed the statistical analysis and led the drafting of this manuscript. JK helped conceive this research; supervised data collection and helped with drafting of this manuscript; CS prepared and aggregated the MTI data and assisted in the drafting of the manuscript; CM, AG, JR, LA and JB all conceived the study

and assisted in the write up of the manuscript. JB is the principal investigator of the study. All authors read and approved the final manuscript.

Disclaimer

The opinions or assertions contained in this manuscript are the private ones of the authors and are not to be construed as official or reflecting the views of NASA or the U.S. Departments of Energy or Defense.

References

1. U.N.: **World Urbanization Prospects: The 2001 Revisions (Key Findings).** *New York: United Nations Population Division* 2001.
2. Robert V, Macintyre K, Keating J, Trape JF, Duchemin JB, Warren M, Beier JC: **Malaria transmission in urban sub-Saharan Africa.** *Am J Trop Med Hyg* 2003, **68:**169-76.
3. Khaemba BM, Mutani A, Bett MK: **Studies of anopheline mosquitoes transmitting malaria in a newly developed highland urban area: a case study of Moi University and its environs.** *East Afr Med J* 1994, **71:**159-164.
4. Knudsen AB, Slooff R: **Vector-borne disease problems in rapid urbanization: new approaches to vector control.** *Bull World Health Organ* 1992, **70:**1-6.
5. Molyneux DH: **Patterns of change in vector-borne diseases.** *Ann Trop Med Parasitol* 1997, **91:**827-39.
6. Trape JF, Zoulani A: **Malaria and urbanization in central Africa: the example of Brazzaville. Part III: Relationships between urbanization and the intensity of malaria transmission.** *Trans R Soc Trop Med Hyg* 1987, **81:**19-25.
7. Keating J, MacIntyre K, Mbogo C, Githeko A, Regens JL, Swalm C, Ndenga B, Steinberg LJ, Kibe L, Githure JI, Beier JC: **A geographic sampling strategy for studying relationships between human activity and malaria vectors in urban Africa.** *Am J Trop Med Hyg* 2003, **68:**357-65.
8. Hay SI: **An overview of remote sensing and geodesy for epidemiology and public health application.** *Adv Parasitol* 2000, **47:**1-35.
9. Hay SI, Omumbo JA, Craig MH, Snow RW: **Earth observation, geographic information systems and *Plasmodium falciparum* malaria in sub-Saharan Africa.** *Adv Parasitol* 2000, **47:**173-215.
10. Beck LR, Lobitz BM, Wood BL: **Remote sensing and human health: new sensors and new opportunities.** *Emerg Infect Dis* 2000, **6:**1-15.
11. Hay SI: **Remote sensing and disease control: past, present and future.** *Trans R Soc Trop Med Hyg* 1997, **91:**105-6.
12. Hay SI, Snow RW, Rogers DJ: **Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data.** *Trans R Soc Trop Med Hyg* 1998, **92:**12-20.
13. Brooker S, Hay SI, Bundy DA: **Tools from ecology: useful for evaluating infection risk models?** *Trends Parasitol* 2002, **18:**70-4.
14. Rogers DJ, Randolph SE, Snow RW, Hay SI: **Satellite imagery in the study and forecast of malaria.** *Nature* 2002, **415:**710-5.
15. Thomson P, Connor S, Milligan P, Flasse S: **Mapping malaria risk in Africa: What can satellite data contribute?** *Parasitol Today* 1997, **13:**313-318.
16. Thomson MC, Connor SJ, Milligan P: **The ecology of malaria – as seen from Earth-observation satellites.** *Ann Trop Med Parasitol* 1996, **90:**243-264.
17. Macintyre K, Keating J, Sosler S, Kibe L, Mbogo CM, Githeko AK, Beier JC: **Examining the determinants of mosquito-avoidance practices in two Kenyan cities.** *Malar J* 2002, **1:**14.
18. Anonymous: **1999 Population and Housing Census: Volume 1.** *Nairobi, Kenya: Republic of Kenya: Central Bureau of Statistics, Ministry of Finance and Planning* 2000.
19. Githeko AK, Service MW, Mbogo CM, Atieli FK, Juma FO: ***Plasmodium falciparum* sporozoite and entomological inoculation rates at the Ahero rice irrigation scheme and the Miwani sugar-belt in western Kenya.** *Ann Trop Med Parasitol* 1993, **87:**379-391.
20. Beier JC, Oster CN, Onyango FK, Bales JD, Sherwood JA, Perkins PV, Chumo DK, Koech DV, Whitmire RE, Roberts CR: ***Plasmodium falciparum* incidence relative to entomologic inoculation rates at a site proposed for testing malaria vaccines in western Kenya.** *Am J Trop Med Hyg* 1994, **50:**529-536.
21. Gimnig JE, Vulule JM, Lo TQ, Kamau L, Kolczak MS, Phillip-Howard PA, Mathenga EM, ter Kuile FO, Nahlen BL, Hightower AW, Hawley WA: **Impact of permethrin-treated bednets on entomological indices in an area of intense year-round malaria transmission.** *Am J Trop Med Hyg* 2003, **68(Suppl 4):**16-22.
22. Mbogo CN, Snow RW, Kabiru EW, Ouma JH, Githure JI, Marsh K, Beier JC: **Low-level *Plasmodium falciparum* transmission and the incidence of severe malaria infections on the Kenyan coast.** *Am J Trop Med Hyg* 1993, **49:**245-253.
23. Service MW: **Mosquito Ecology.** *New York: John Whiley and Sons Publishing* 1976.
24. Kay RA, Bender SC, Henson TD, Byrd DA, Rienstra JL, Decker ML, Rackley NG, Akau RL, Claassen PJ, Kidner RE, Rtaplin RB, Bullington DM, Marbach KD, Lanes CE, Little C, Smith BW, Brock BC, Weber PG: **Multispectral thermal imager (MTI) payload overview.** In: *SPIE Annual Conference: Technical Conference 3753; Denver, Colorado.* 1999.
25. Kalton G: **Introduction to Survey Sampling.** *Newbury Park: Sage Publications* 1983.
26. Grimm NB, Grove M, Pickett STA, Redman CL: **Integrated approaches to long-term studies of urban ecological systems.** *BioScience* 2000, **50:**571-584.
27. Chinery WA: **Impact of rapid urbanization on mosquitoes and their disease transmission potential in Accra and Tema, Ghana.** *East Afr Med J* 1995, **2:**179-188.