ArcGIS and Malaria: Evaluating environmental effects on mosquito distributions in Tanzania using GIS, Remote Sensing, and Python

By: Emily Acheson
Department of Biology, University of Ottawa

NSERC CRSNG

Application for the 2015 ESRI Canada Scholarship Award

BACKGROUND

Sub-Saharan Africa continues to be ravaged by *Plasmodium falciparum* malaria, a debilitating and often fatal disease carried by its disease vector: *Anopheles* mosquitoes (Kelly-Hope *et al.*, 2009). In Tanzania, one of Africa's most affected countries, malaria still causes the highest morbidity and mortality rates (Mboera *et al.* 2011).

Increasingly sophisticated statistical models are being developed to predict disease vector distributions (e.g. Kulkarni *et al.*, 2010), where studies on the spatiotemporal patterns of these vectors and their diseases is urgently needed (Siraj *et al.*, 2014). These models cannot be built without GIS and remote sensing tools.

Although GIS and remote sensing have been used extensively to analyze satellite images of environmental factors (e.g. elevation, temperature, etc.) for disease vector research, these tools have never before been combined to study the effects of disease vector controls.

For instance, Tanzania conducted a country-wide distribution of insecticide-treated bed nets across Tanzania in 2009 (Kabula et al. 2011). How do mosquito nets affect mosquito distributions across Tanzania? I attempt a new approach to GIS analysis by using bed net data with mosquito habitat suitability.

Our study investigates the driving forces for changes in the geographical distribution of *Anopheles* mosquitoes in country-wide Tanzania, including mosquito net distributions, between 2001 and 2011. I therefore deal with the known and possible drivers of geographical change in these malaria vectors.

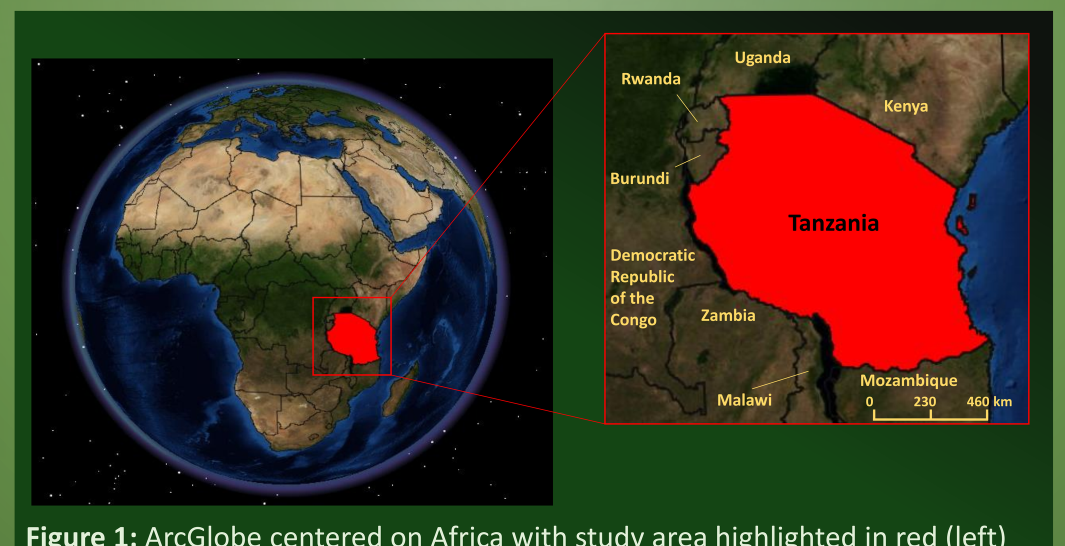

Figure 1: ArcGlobe centered on Africa with study area highlighted in red (left) with zoomed-in focus on study area and surrounding countries (right).

Biological questions:

1) Do mosquito nets decrease mosquito habitat suitability?
2) Are bed nets being used optimally based on mosquito habitat suitability (i.e. is mosquito net intensity greater in areas of greater habitat suitability)?

If insecticide-treated mosquito nets are effective not only as a physical barrier to prevent mosquitoes from biting humans but also as a chemical barrier to kill mosquitoes, I predict mosquito nets will decrease mosquito habitat suitability. I also predict that intensity of mosquito net use will increase as mosquito habitat suitability increases.

SOURCES

Araujo *et al.*, "Reducing uncertainty in projections of extinction risk from climate change." *Global Ecology and Biogeography*. **14** (2005).

Kabula *et al.*, "Malaria entomological profile in Tanzania from 1950-2010: A review of mosquito distribution, vectorial capacity and insecticide resistance." *Tanzania Journal of Health Research*. **13** (2011).

Kelly-Hope *et al.*, "Environmental factors associated with the malaria vectors Anopheles gambiae and Anopheles funestus in Kenya." *Malaria Journal*. **8** (2009).

Kulkarni *et al.* "High Resolution Niche Models of Malaria Vectors in Northern Tanzania: A New Capacity to Predict Malaria Risk?" *PlosOne*. (2010).

Mboera, *et al.* "Impact of climate change on human health and health systems in Tanzania: a review." *Tanzania Journal of Health Research*. (2011).

Phillips *et al.* "Maxent entropy modelling of species geographic distributions." *Ecological Modeling.* (2006).

Siraj *et al.* "Altitudinal Changes in Malaria Incidence in Highlands of Ethiopia and Colombia." *Science*. **343** (2014).

METHODS

Modelling mosquito habitat suitability

To determine mosquito habitat suitability, we used the Maxent program (Version 3.3.3k, Phillips et al., 2006), a species habitat modelling software. Maxent correlates input variables (e.g. temperature) with the georeferenced occurrence locations of a species to produce a model that predicts the range of the species over a given area. For our environmental inputs, we used the following (Figure 2):

A) Average Land Surface Temperature
B) Average Normalized Difference Vegetation Index (NDVI)
C) Human Population Density
D) Land Cover
E) Elevation

These environmental inputs were acquired for 2001 and for 2011 from the NASA satellite MODIS (Moderate-Resolution Imaging Spectroradiometer). We compare models in 2001 and 2011 due to environmental data availability from MODIS, which starts in 2001. NDVI acted as a proxy for precipitation, since MODIS does not measure precipitation directly.

Python scripts were used in ArcGIS 10.1 to mosaic each Environmental layer, match their projections, clip to the extent of Tanzania, and convert all layers to ASCII format before they were imported into Maxent for model building.

Anopheles mosquito occurrence records were collected from over 400 published articles. For the 2001 model, occurrence records were used from 1999 to 2003 (n=56), and for the 2011 model, records from 2009 to 2013 were used (n=83).

Within Maxent, each model was run with 10-fold crossvalidation. The averaged model and corresponding AUC (area under the curve) value of spatial accuracy were assessed. Each input variable was ranked based on the importance of its contribution to building the model.

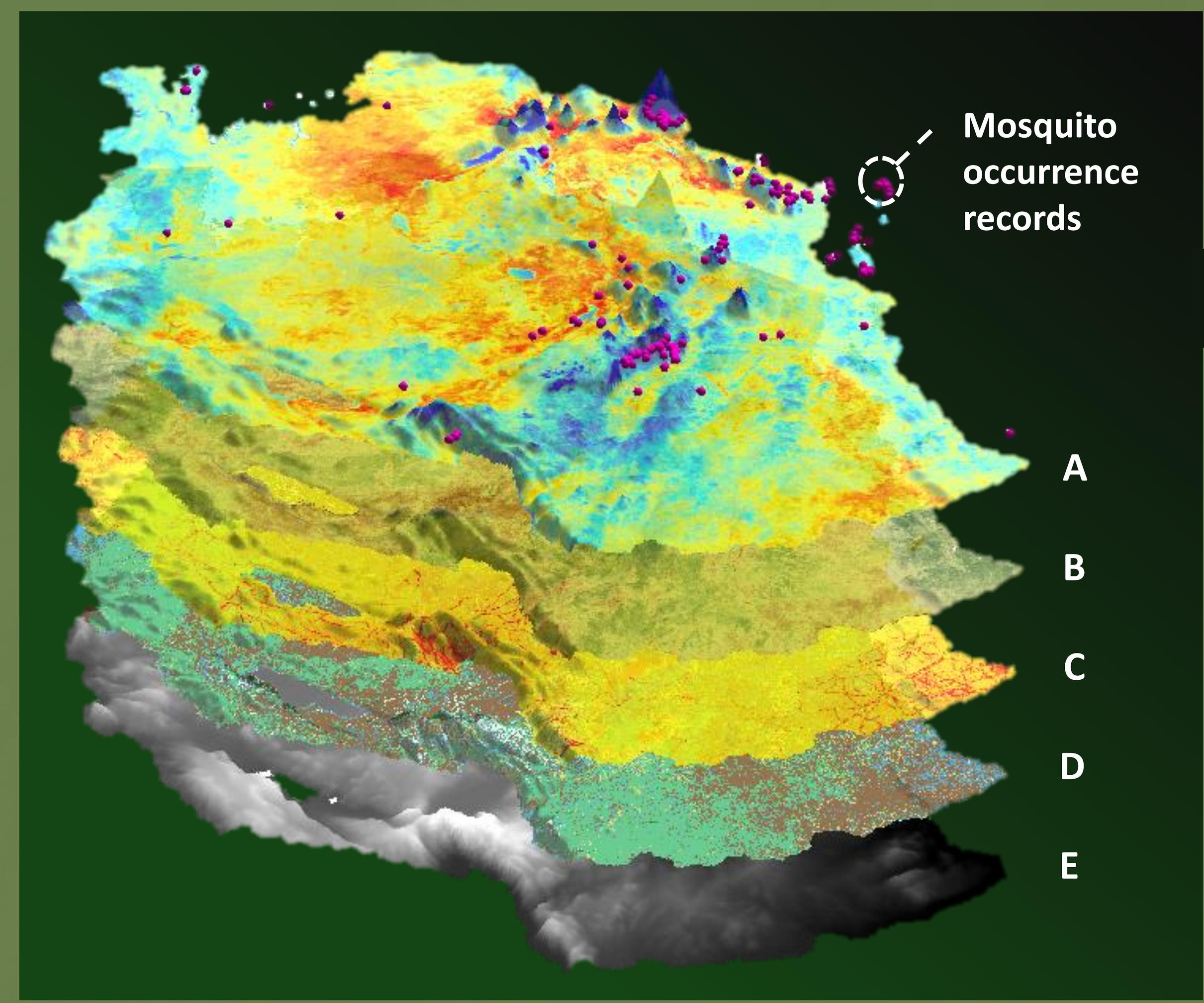

Figure 2: A 3-D depiction of the five environmental input layers used to build the 2001 Maxent model (the corresponding 2011 variables were used for the 2011 model). The mosquito occurrence records, shown as purple spheres, are also shown. This figure is original and was created in ArcScene 10.1.

Creating a spatial depiction of mosquito net usage

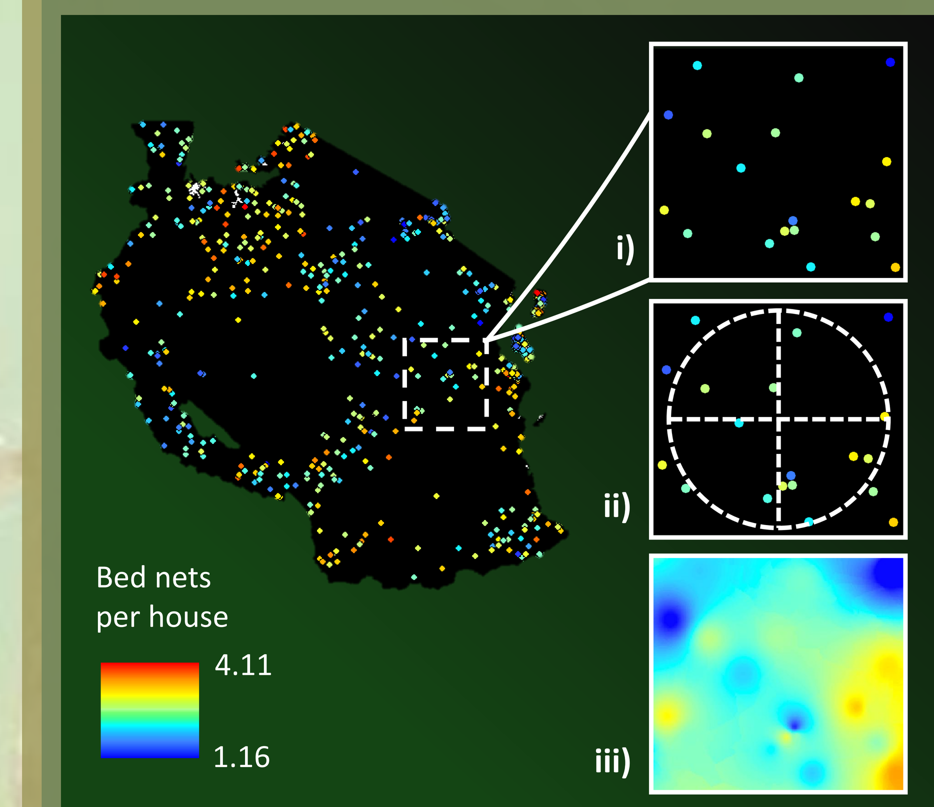

Figure 3: Process to create an interpolated bed net layer from i) point locations of bed net use across Tanzania (left), where ii) 12 neighbouring points are used to calculate each pixel to create iii) the interpolated surface.

The georeferenced mosquito net usage was provided by the Demographic Health Surveys (DHS), with permission. The 573 point locations were given across Tanzania, with each point representing clusters of approximately 18 nearby households and each household's bed net use, including the number of bed nets used per house.

The following steps were taken to create an interpolated layer of bed net use (Figure 3):

Step 1: Georeferenced cluster points were imported into ArcGIS 10.1 and processed for projection and spatial extent (Figure 3i).

Step 2: Inverse-distance weighting (IDW) interpolation was conducted in ArcGIS due to the clustering of the points. Twelve neighbouring points were used for interpolation of each pixel (Figure 3ii), with neighbouring points losing weight as distance increases. Interpolation with a power of 2 was used.

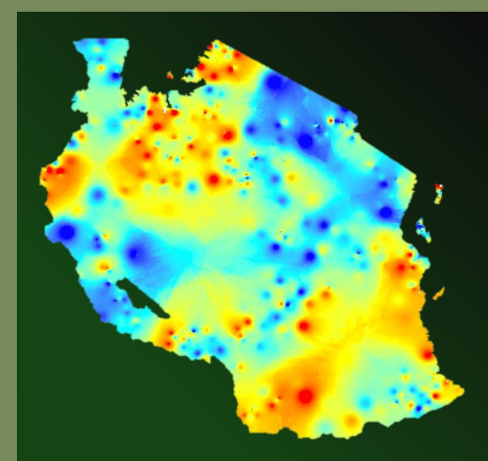

Figure 4: Final interpolated layer after crossvalidation.

Step 3: Once the interpolated layer was created (Figure 4), I split the original bed net data set for training (90%) and testing (10%) and conducted 10-fold crossvalidation. Using Python, I created 10 IDW layers. I imported these data into the R statistics program (Version 0.98.1091) and validated each layer with its associated testing data.

RESULTS

Maxent was run on the five input layers to create predictive models and maps of *Anopheles* mosquito habitat suitability for 2001 and 2011 (maps shown in Figure 5). The habitat suitability maps represent the relative suitability through space. Both models rank elevation, human population, and land cover as the top three factors for building each model (with a combined contribution of 90%). AUC values were high (0.86 for 2001 and 0.87 for 2011). However, the 2001 model suggests a more extensive area of mosquito habitat suitability relative to that of 2011.

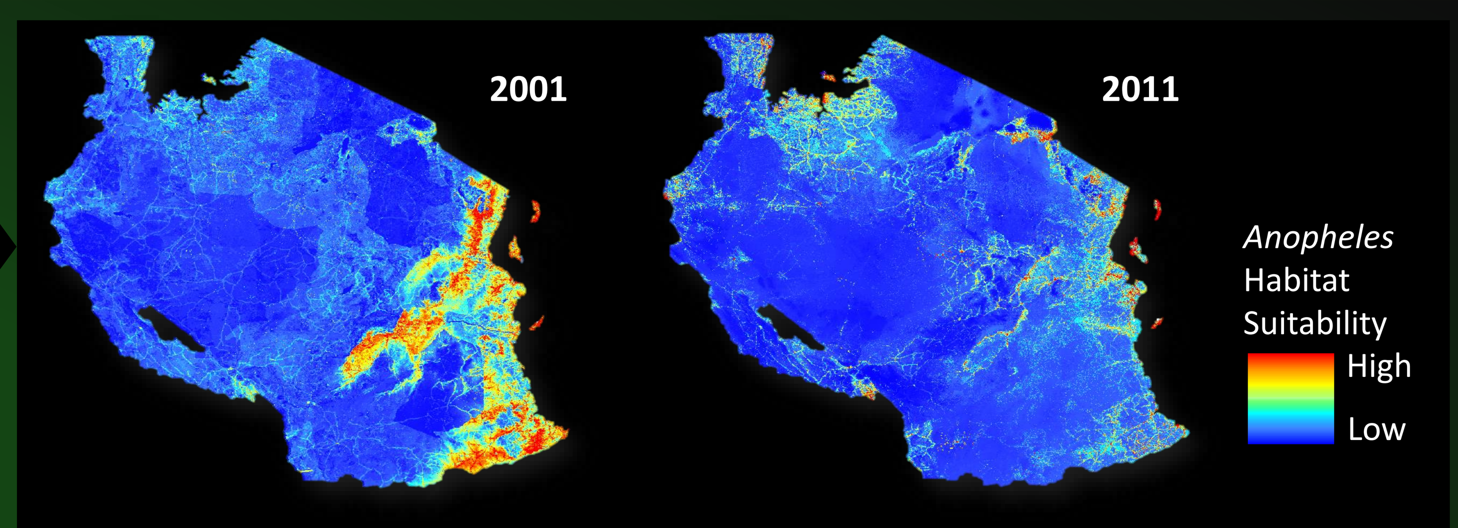

FIGURE 5: Maxent output of predicted habitat suitability for *Anopheles* mosquitoes in Tanzania in 2001 and 2011.

Are mosquito nets responsible for this apparent decrease in mosquito habitat suitability?

In order to explore this further, I imported the Maxent outputs into ArcGIS 10.1 and used the "Raster Calculator" to subtract the 2001 habitat suitability map from that of 2011. The resulting layer, representing the change in predicted *Anopheles* habitat suitability, suggests that mosquito habitat suitability has remained relatively unchanged between 2001 and 2011, but that suitability has increased across many northern areas and decreased along low elevation areas along the east coast (Figure 6).

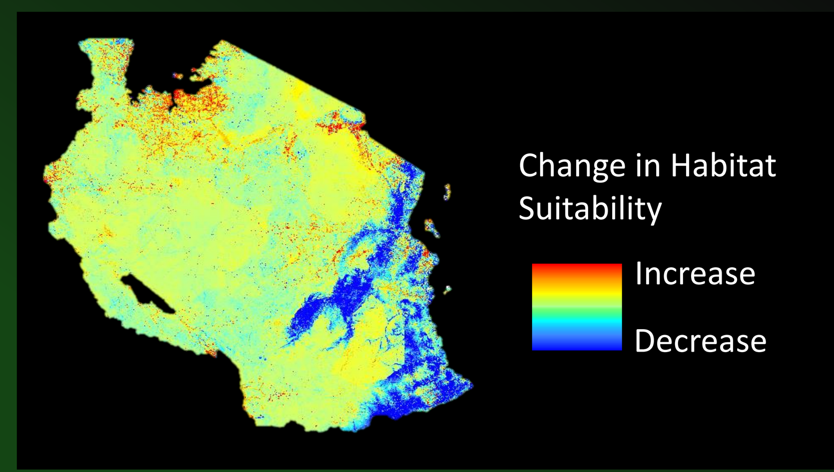

FIGURE 6: Change in *Anopheles* habitat suitability between 2001 and 2011.

With a layer of change in mosquito habitat suitability and a layer of interpolated bed net use across country-wide Tanzania, I can now import these layers into R to compare the two layers. Since there are over 1 million pixels, I randomly selected 500 points for easier visual representation (Figure 7).

The resulting correlation analysis gives the following regression equation:

$y = -0.03147x + 0.11111$

Though there is a negative correlation between bed net use and change in habitat suitability, the associated R^2 value is 0.00414 and a p-value of 0.1286.

Correlate interpolated bed net layer with change in mosquito habitat suitability layer

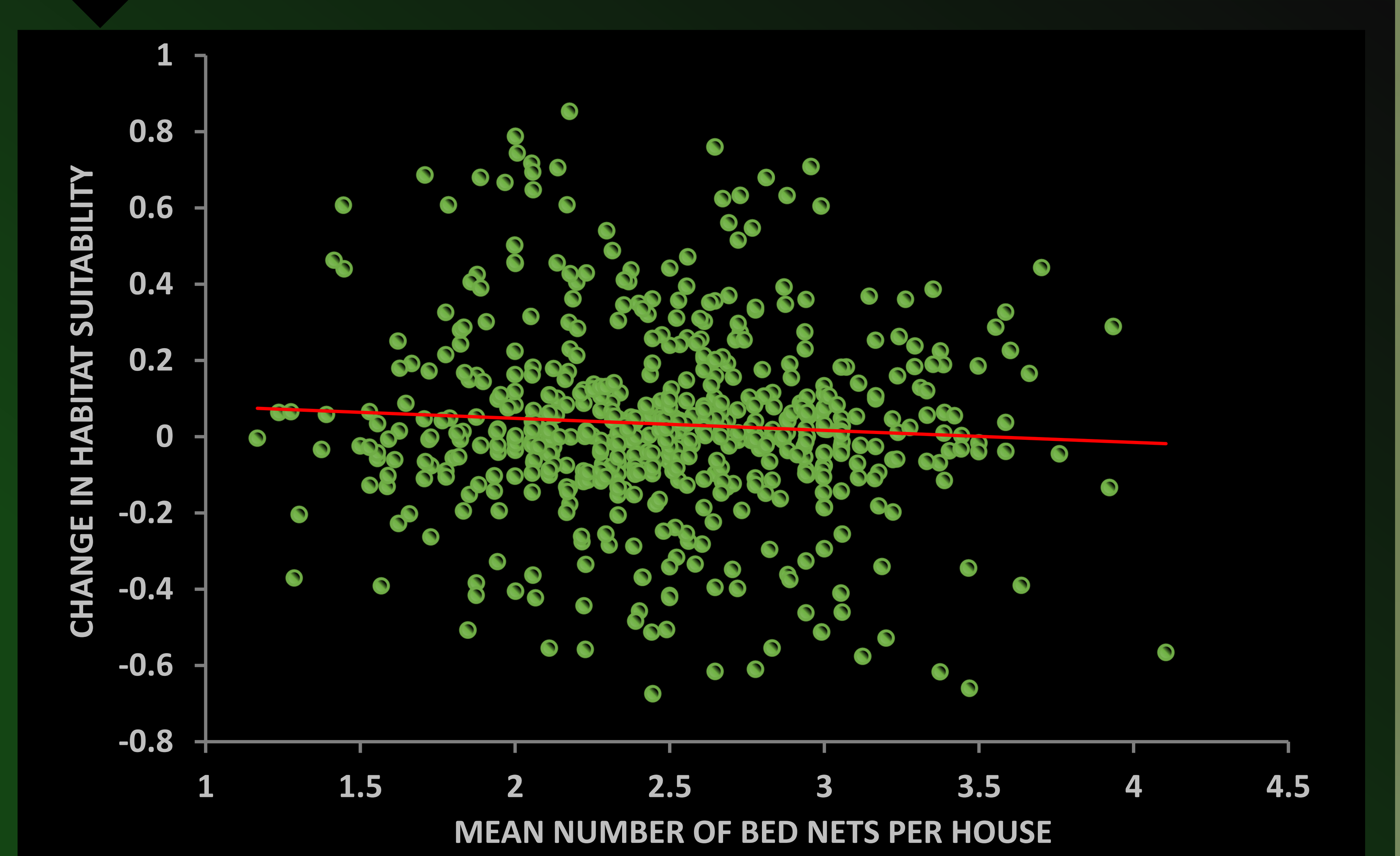

FIGURE 7: Dot plot with trend line (in red) representing the association between the average number of bed nets per house (i.e. mosquito net intensity). A random 500 pixels were selected for testing.

CONCLUSIONS AND FUTURE STEPS

1) Do mosquito nets decrease mosquito habitat suitability?

Based on the results shown in Figure 7, *Anopheles* mosquito habitat suitability does not appear to be correlated with the countrywide distribution of insecticide-treated bed nets in 2009. The low R^2 value and p-value that exceeded the significance threshold of 0.05 allow me to conclude that the mosquito nets do not decrease mosquito habitat suitability.

2) Are bed nets being used optimally based on mosquito habitat suitability?

The results of Figure 7 are one way to approach this question. Since no correlation was found between mosquito habitat suitability and bed net intensity, I can conclude that bed nets are not being used optimally.

Areas with increases in mosquito habitat suitability should have greater bed net usage than areas with low habitat suitability. Currently, bed net use across countrywide Tanzania is nearly random.

What are future steps for this project and how will ArcGIS and Python be involved?

This research project represents the first known attempt to combine vector control programs with species distribution modelling. ArcGIS and Python are the vital tools necessary to make this kind of research possible by processing the spatial data.

Bed net use in Tanzania must be optimized to target areas of moderate to high *Anopheles* habitat suitability to protect people from malaria and save lives.

Another way to approach the above biological questions is to focus on certain areas of Tanzania where bed nets are clustered and less interpolation is needed. This will allow a separate, and potentially more accurate, comparison of bed net use and habitat suitability.

This research is crucial to informing expensive control programs and practices on how best to implement their strategies. Models can also be validated through time and projected into the future under changing environmental scenarios to predict potential future changes in mosquito habitats and to determine how to optimize bed net use.