

# Understanding the social and environmental drivers of malaria – a current risk map for Ghana.

## Project proposal

Although there have been large efforts to map the malaria risk and its distribution spatially and temporally globally (Gething et al., 2011) they are built from a historical framework of spatio-temporal data, that creates less accurate estimates of the current situation of disease risk. Since the birth of Roll Back Malaria in 2000, surveillance of malaria has been increased giving rise to several large datasets such as the Demographic Health Surveys (DHS) and in Malaria Indicator Surveys (MIS). There are currently no published maps for Ghana's current malaria situation even though in 2011, UNICEF in collaboration with the National Malaria Control Programme was able to implement a national Malaria Indicator Cluster Survey (MICS), which includes high resolution data on key malaria indicators including prevalence and incidence (PMI, 2015). A current risk map would help guide a more targetted approach to malaria control for the National Malaria Control Program and allow optimised resource allocations.

My aim is to combine different large-scale datasets in an innovative geospatial models to create improved risk maps with greater accuracy. The integration of databases allows for higher spatial resolution providing less uncertainty in prediction estimates. Currently the most recent large-scale datasets available are the Demographic Health Survey (Ghana Statistical Service & Ghana Demographic Health Survey, 2014) and Multiple Indicator Cluster Survey by UNICEF ("UNICEF: Multiple Indicator Cluster Surveys: Available reports/datasets. [[http://www.childinfo.org/mics%7B\\_%7Davailable.html](http://www.childinfo.org/mics%7B_%7Davailable.html)],” n.d.). However, given the difference in collection methods, these data will have to be carefully processed and weighted in order to reduce collection bias for further modelling purposes. Other datasets that would be required would be the extraction of spatial rasters that can be linked to survey clusters for the geospatial models. These would be sourced from open GIS sources such as worldclim, NASA, NOAA etc.

### Reference:

1. Gething, P. W., Patil, A. P., Smith, D. L., Guerra, C. A., Elyazar, I. R. F., Johnston, G. L., . . . Hay, S. I. (2011). A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria Journal*, 10, 378. doi:10.1186/1475-2875-10-378
2. Ghana Statistical Service, & Ghana Demographic Health Survey. (2014). Ghana Demographic and Health Survey 2014: Ghana Statistical Service, Ghana Health Service, Ghana AIDS Commission. DHS. Retrieved from <https://dhsprogram.com/publications/publication-FR307-DHS-Final-Reports.cfm>
3. PMI. (2015). President's Malaria Initiative Ghana Malaria Operational Plan FY 2015. Retrieved from <http://www.pmi.gov/docs/default-source/default-document-library/malaria-operational-plans/fy-15/fy-2015-ghana-malaria-operational-plan.pdf?sfvrsn=3>
4. UNICEF: Multiple Indicator Cluster Surveys: Available reports/datasets. [[http://www.childinfo.org/mics%7B\\_%7Davailable.html](http://www.childinfo.org/mics%7B_%7Davailable.html)]. (n.d.). Retrieved from [http://www.childinfo.org/mics\\_available.html](http://www.childinfo.org/mics_available.html)

## What has been done:

Given the difficulty with automating the downloading of remote sensed imagery from MODIS, only data cleaning and exploratory analysis has been conducted. These are important steps mainly to try and understand the data. The datasets are different in terms of their collection and how they code their variables. The most important variables in our case is Microscopy results for malaria. RDT results have also been captured for potential future work. Other response variables also considered are Hemoglobin, fever and malaria based severe Anaemia (defined by WHO as hemoglobin below 7.0 g/dL).

## What is available:

NOTE: Data is too large to upload onto gitHub and so the Rproject will be directly sent to you. It has been zipped and will need unzipping.

Data from MICS and DHS have been imported and stored in data files in an Project file. with file path:

- i. for MICS (Data/Original/MICS/mics4 datasets)
- ii. for DHS (Data/Original/DHS)

this data includes, household data, child data, laboratory data (for MICS) and gps data. All original file names have been kept incase of new versions (new version expected to be released for DHS in January). Data has been cleaned and ready for analysis. As part of data cleaning some exploratory analysis was conducted for MICS and DHS separately (see sections names exploratory analysis and GIS analysis)

GIS cleaning for MICS 2011: allows for importing raster layers and cropping, re-projecting and aligning (by extent) for data extraction. A final dataset ready for geospatial modeling for MICS 2011 has been made (found in folder named Data/Final\_for\_analysis).

A short exploratory analysis and GIS analysis for each dataset (please see attached short report).

## Steps to go through:

### Open Script labeled Data\_cleaning.r

1. This script first will take you through MICS 2011.

#### I. Data cleaning

- Importing of all datasets (includes Household members, laboratory tests, individual child and gps data)
- Data is then cleaned to only keep important variables for malaria as defined by a literature review conducted outside the scope of this project (in press Millar et al. 2017).
- the data is then attached to gps data for further analysis.
- Lastly, important variables such as age, microscopy, RDT results and hemoglobin were recoded, this is because the dataset labeled missing data differently or to convert string data to numeric binary 0 or 1.

#### II. Exploratory Analysis

- Before running into prediction modeling it would be useful to determine if key variables that are identified in the literature are strongly related to response variable (Microscopy in this case).
- We ran plots to see how Age interacted with RDT and Microscopy
- Plots to see how wealth and residence are distributed by microscopy results
- Boxplot to see if there is a strong relationship between hemoglobin levels and microscopy and to determine in general if positive microscopy is linked to lower hemoglobin levels (i.e. severe anaemia).

#### III. GIS Analysis

- In GIS analysis the main aim was to plot the point prevalence of MICS 2011 to see the general geographic layout of prevalence across the country in 2011.

2. The script then goes through the section for DHS 2014:

#### I. Data cleaning

- Importing of all datasets (includes Household members, individual child and gps data).
- Data is then cleaned to only keep important variables for malaria as defined by a literature review conducted outside the scope of this project (in press Millar et al. 2017).

- the data is then attached to gps data for further analysis.
- Lastly, important variables such as microscopy, RDT results and hemoglobin were recoded, this is because the dataset labeled missing data differently or to convert string data to numeric binary 0 or 1.

## II. Exploratory Analysis

- Before running into prediction modeling it would be useful to determine if key variables that are identified in the literature are strongly related to response variable (Microscopy in this case).
- We ran plots to see how Age interacted with RDT and Microscopy
- Plots to see how wealth and residence are distributed by microscopy results
- Boxplot to see if there is a strong relationship between hemoglobin levels and microscopy and to determine in general if positive microscopy is linked to lower hemoglobin levels (i.e. severe anaemia).

## III. GIS Analysis

- In GIS analysis the main aim was to plot the point prevalence of DHS 2014 to see the general geographic layout of prevalence across the country in 2014 and compare to trends from 2011.

### Open Script labeled GIS data cleaning.r

1. This script is mainly to clean raster layers, by means of cropping, aligning extents and re-projecting. Note that only 2011 variables have been collected so far. This is owing to a set back with automating the process of downloading MODIS tiles.
  - Import all un-processed rasters. These are directly taken from web sources such as MODIS, VIIRS etc.
  - Import shape file for Ghana (used for cropping)
  - Send all rasters through a for loop that crops, aligns and reprojects them all to the same coordinate system (a template raster is used). The projection is WGS1984 in Long and Lat format. The for loop exports the rasters to a file that stores all cleaned rasters (NOTE: this for loop takes a while, It has been pre-run for you to observe results). the rasters are stacked and plotted to visualize the stack.
  - finally the MICS data with gps coordinates are brought in, converted to a shape file and the raster values are extracted for each location.
  - For analysis, the data must be standardized, they are scaled and exported as a .csv file to a final folder that contained ready to use .csv files for use in the geospatial models.

## MODIS Downloading scripts

These include:

1. mainDownload.R (run first)
2. modisDownload.R (has all the functions)
3. modisDownload.json (includes file directories and tile + product information for downloading)
4. ModisLP.Rdata (empty Rdata for storing temporary information)
5. agroclimate.netrc (login and password file - make sure to input your own!)

NOTE: these are not for sharing and have been produced by the *AgroClimate* team at University of Florida, Department of Agriculture and Biological Engineering. However, I have included them educationally and incase the scripts seem interesting to those invited to this gitHub.

NOTE: be sure to change all file directories appropriately for it to work. The connection can be annoying and so far has not been able to download more than two tiles at a time. Also for these scripts to work you will need to download the *MRT Tool*