Malaria Project

Brendon Cornett

October 26, 2018

Background and Introduction

Malaria research is important. Data research into malaria spread and its relationship to available variables is relatively new. With the kinds and amount of information available today, the creation of infectivity algorithms, climate models and treatment efficacy are all possible. While at the same time, biological controls based on these models are being implemented at record pace. There is great difficulty with Malarial control given the unique nature of the disease. It has a long lag time for symptoms, requires a complex life cycle to propogate giving a large amount of variablility that has only recently been analyzed given the computational intensity of such a task. With all this known, even advanced models tend to break down when applied to real life situations, and a researcher can spend a lifetime modeling the unique situation of malaria spread in a single region.

Unfortunately, even with the controls in place, the state of the art science today, erradication of malaria seems to be but a pipedream given the financial investment it would take. Thus we are stuck using these imperfect models to attempt some measure of control. The simple analysis taking place in this document is far from comprehensive enough to take seriously in any true use scenario, but rather exists as a short exploratory template for interacting with similar malaria data.

Malaria simply requires three things to occur, mosquitoes, humans, and the parasite.[1] Humans are found everywhere, but in places with greater density of people, better propagation of the parasite is NOT expected. Indeed, humans in rural environs find a greater risk of malaria.[3] The Mosquito is the more complex actor in the equation of malaria. It has very specific requirements for its breeding. It is directly linked to rainfall and temperature linearly. More rain or higher temperature induces greater mosquito breeding.[2] Thus for malaria to exist, a tropical or subtropical climate is need. Given this, understanding the rainfall and temperature of your country, region, or other units of division can be your greatest tool in understanding malaria risk and prevalence. While this document will not be exhaustive, it will attempt to explain this relationship using Mozambique as representative of countries dealing with Malaria on a national scale. In particular there is a greater risk of damage for youths, those youngest are most vulnerable.[1]

Understanding Rain and Temperature

Given the appeal of using climate to approximate malaria ripe conditions wetter hotter areas of the country would see more malaria. Regionally, one can see below that the Costal region has the greatest raw amount of malaria cases over the 2010 to 2016 period. Although the other regions are also quite close with the exception of the Southern region.

```
## # A tibble: 4 x 2
##   Region   `sum(cpt)`
##   <fct>         <dbl>
## 1 Center      131880.
## 2 Coastal     219173.
## 3 Northern    211552.
## 4 Southern     31535.
```

The average weekly temperature of the different regions is not all that different, nor is the average weekly rainfall totals.

```
## # A tibble: 4 x 2
##   Region   `mean(tavg)`
##   <fct>           <dbl>
## 1 Center           22.5
## 2 Coastal          24.1
## 3 Northern         22.8
## 4 Southern         23.3
```

```
## # A tibble: 4 x 2
##   Region   `mean(rainTot)`
##   <fct>              <dbl>
## 1 Center              7.25
## 2 Coastal             6.17
## 3 Northern            6.46
## 4 Southern            5.98
```

With this data it could be assumed everything discussed in this document was wrong then about malarial association with climate! Obviously, this is incorrect, there must be something else effecting malaria in the south. Perhaps the variation in climate is greater across the southern region?

```
## # A tibble: 4 x 2
##   Region   `var(cpt)`
##   <fct>         <dbl>
## 1 Center         28.6
## 2 Coastal        44.0
## 3 Northern       77.6
## 4 Southern       33.7
```

```
## # A tibble: 4 x 2
##   Region   `var(tavg)`
##   <fct>          <dbl>
## 1 Center         13.7
## 2 Coastal         7.30
## 3 Northern        8.72
## 4 Southern        9.54
```

```
## # A tibble: 4 x 2
##   Region   `var(rainTot)`
##   <fct>             <dbl>
## 1 Center             362.
## 2 Coastal            408.
## 3 Northern           336.
## 4 Southern           358.
```

That's not it. In fact, the central and coastal regions have greater variation than the others in terms of average temperature and weekly total rainfall respectively.

Perhaps then it is not the average temperature, but its extremes?

```
## # A tibble: 4 x 2
##   Region   `sum(tabove30)`
##   <fct>              <int>
## 1 Center              2929
## 2 Coastal             1237
## 3 Northern             171
## 4 Southern             794
```

```
## # A tibble: 4 x 2
##   Region   `sum(tabove35)`
##   <fct>              <int>
## 1 Center                36
## 2 Coastal                4
## 3 Northern               0
## 4 Southern               1
```

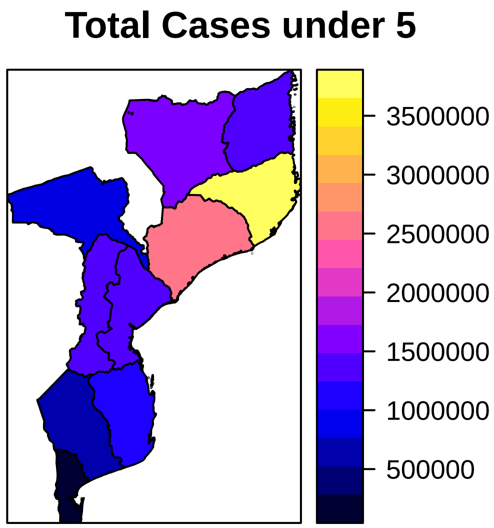
```
## # A tibble: 4 x 2
##   Region   `sum(tbelow15)`
##   <fct>              <int>
## 1 Center              2474
## 2 Coastal               46
## 3 Northern            1158
## 4 Southern              94
```

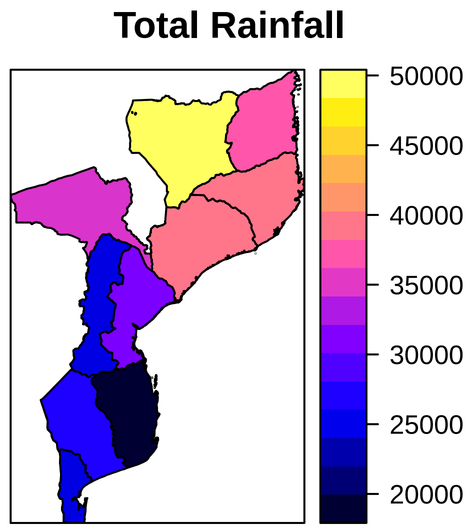
Here is something one can see as being quite distinct among the regions with high malaria incidence. Those regions with the hotter recorded temperatures are the very same as those regions with higher malarial incidence! While at the same time these same regions have higher numbers of low temperature days. Thus, ironically looking directly at the variance gives the false impression of uniformity.

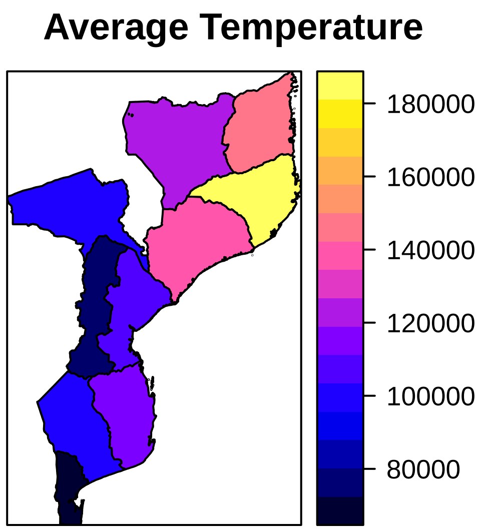
Also effecting malaria occurrence is the importance of the Southern region's demographics. It is by large the most urbanized region, and thus the region least prone to mosquitoes regardless of the climate. With this in mind, one can see that rurality and its economics is a most important factor influencing malaria exposure and spread.

As an aside this is the main reason why, despite having the correct climate for it, first world nations like the U.S. have largely eradicated their malaria. The healthcare and urbanization of America means regions with the potential are well kempt, and those that might have exposure are given adequate access to quinine.

Could it then be that looking at the regions themselves are too large? If we both zoom out and zoom in, looking at malarial cases at the province level we can see something very surprising about the north that cannot be seen from the numbers alone.



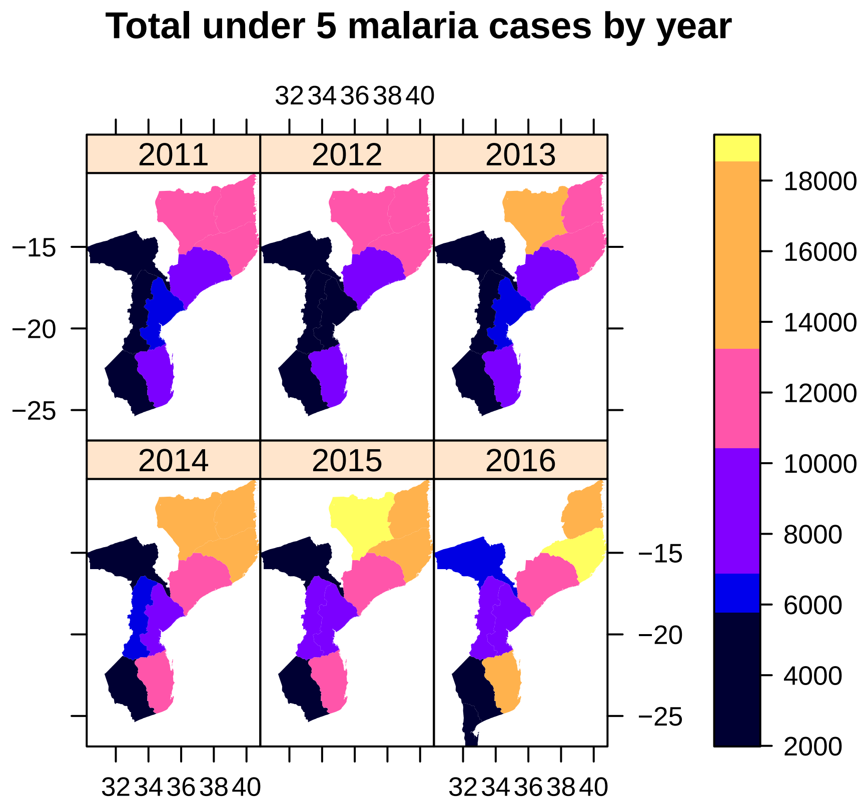




This looks pretty good now. We can see that in the northern region there is a large amount of malaria AND high average temperature. This province is driving up the numbers when we looked at the malaria data purely from a regional standpoint.

Changes in Malaria over time

So far we have only looked at the differences in malaria incidence as it pertains to the region as a whole, combining the years. But what if we look at the yearly data? Is incidence changing over the time?



Looking at this plotting of cases by year, its noticable that malaria incidence is worsening over time across much of Mozambique. The south stays relatively fine over time, but the northern region in particular looks as though every year the incidence rises. This would be consistant with the outlook forcast the UN preformed on Malaria rates as related to climate change.

Understand when Rainfall and Temperature effect Malaria

Obviously malaria does not instantly strike after a good rainstorm, nor does a hot day lead to increases immediately either. Given that the mosquito life stages from egg to adult take between 7 to 14 days, one can guess that there is an increase in malaria 1 to 2 weeks after a good storm or heatwave. Given that this data begins when an individual enters the clinc system, how long was it since the climate event that caused the increase in malaria?

Average Temperature

Looking at the complicated numbers that generate lags give us Average Temperature using the Poisson model at 8 weeks. Thus malaria lags behind Average Temperature by 8 weeks. Looking at the numbers the Means Square Error of 8 weeks was 46.4698622 which was lower than that of the 2 week lag at 47.6218643 or 4 week lag: 47.5313301.

Total Rainfall

Looking at the complicated numbers that generate lags give us Total Rainfall using the Poisson model at 2 weeks. Thus malaria lags behind total rainfall by 2 weeks. Looking at the numbers the Means Square Error of 2 weeks was 46.3644015 which was lower than that of the 4 week lag at 45.4612548 or 8 week lag: 45.6786648.

Days above 30C

Looking at the complicated numbers that generate lags give us Days above 30C using the Poisson model at 8 weeks. Thus malaria lags behind Days above 30C by 8 weeks. Looking at the numbers the Means Square Error of 8 weeks was 46.4698622 which was lower than that of the 2 week lag at 47.6218643 or 4 week lag: 47.5313301.

Days above 35C

Looking at the complicated numbers that generate lags give us Days above 35C using the Poisson model at 2 weeks. Thus malaria lags behind Days above 35C by 2 weeks. Looking at the numbers the Means Square Error of 2 weeks was 47.9724572 which was lower than that of the 4 week lag at 48.0670602 or 8 week lag: 48.2545635.

Days below 15C

Looking at the complicated numbers that generate lags give us Days below 15C using the Poisson model at 4 weeks. Thus malaria lags behind Days below 15C by 4 weeks. Looking at the numbers the Means Square Error of 4 weeks was 47.4903439 which was lower than that of the 2 week lag at 47.5705684 or 8 week lag: 47.8052154.

Days below 20C

Looking at the complicated numbers that generate lags give us Days below 20C using the Poisson model at 2 weeks. Thus malaria lags behind Days below 20C by 2 weeks. Looking at the numbers the Means Square Error of 2 weeks was 47.1406492 which was lower than that of the 4 week lag at 46.6130033 or 8 week lag: 46.3407634.

Overall, the Total Rainfall lagging at 2 weeks is the most promising in explaining the child malaria incidence.

References

1. Ecology of Malaria. CDC. https://www.cdc.gov/malaria/about/biology/ecology.html
2. Anopheles quadrimaculatus. University of Florida. http://entomology.ifas.ufl.edu/creatures/aquatic/Anopheles_quadrimaculatus.htm
3. Kabiara et al. The impact of urbanization and population density on childhood Plasmodium falciparum parasite prevalence rates in Africa. Malaria Journal. 2017. https://doi.org/10.1186/s12936-017-1694-2

Backend Code

```
library(tidyverse)
library(modelr)
library(sp)
library(maptools)
library(lattice)
library(latticeExtra)
library(rgdal)
library(RColorBrewer)
library(sf)
library(classInt)
library(forcats)
library(modelr)

# Data creation for lag times

malload = read.csv("./MozSyntheticMalaria.csv")

mozmal = subset(malload, Epiyear != 2017)

mozmal$cpt = (mozmal$malaria/(mozmal$Population_UN * mozmal$u5weight)) *
    1000

mal2W <- mal4W <- mal8W <- mozmal

mal2W = mal2W %>% arrange(Epiyear) %>% group_by(District) %>%
    mutate(rainLAG2W = lag(rain, 2, order_by = District), rainTotLAG2W = lag(rainTot,
        2, order_by = District), tavgLAG2W = lag(tavg, 2, order_by = District),
        rhLAG2W = lag(rh, 2, order_by = District), sdLAG2W = lag(sd,
            2, order_by = District), psfcLAG2W = lag(psfc, 2,
            order_by = District), tabove30LAG2W = lag(tabove30,
            2, order_by = District), tabove35LAG2W = lag(tabove35,
            2, order_by = District), tbelow20LAG2W = lag(tbelow20,
            2, order_by = District), tbelow15LAG2W = lag(tbelow15,
            2, order_by = District), pabove1LAG2W = lag(pabove1,
            2, order_by = District), pabove50LAG2W = lag(pabove50,
            2, order_by = District), pabove100LAG2W = lag(pabove100,
            2, order_by = District)) %>% drop_na()

mal4W = mal4W %>% arrange(Epiyear) %>% group_by(District) %>%
    mutate(rainLAG4W = lag(rain, 4, order_by = District), rainTotLAG4W = lag(rainTot,
        4, order_by = District), tavgLAG4W = lag(tavg, 4, order_by = District),
        rhLAG4W = lag(rh, 4, order_by = District), sdLAG4W = lag(sd,
            4, order_by = District), psfcLAG4W = lag(psfc, 4,
            order_by = District), tabove30LAG4W = lag(tabove30,
```

```
            4, order_by = District), tabove35LAG4W = lag(tabove35,
            4, order_by = District), tbelow20LAG4W = lag(tbelow20,
            4, order_by = District), tbelow15LAG4W = lag(tbelow15,
            4, order_by = District), pabove1LAG4W = lag(pabove1,
            4, order_by = District), pabove50LAG4W = lag(pabove50,
            4, order_by = District), pabove100LAG4W = lag(pabove100,
            4, order_by = District)) %>% drop_na()

mal8W = mal8W %>% arrange(Epiyear) %>% group_by(District) %>%
    mutate(rainLAG8W = lag(rain, 8, order_by = District), rainTotLAG8W = lag(rainTot,
        8, order_by = District), tavgLAG8W = lag(tavg, 8, order_by = District),
        rhLAG8W = lag(rh, 8, order_by = District), sdLAG8W = lag(sd,
            8, order_by = District), psfcLAG8W = lag(psfc, 8,
            order_by = District), tabove30LAG8W = lag(tabove30,
            8, order_by = District), tabove35LAG8W = lag(tabove35,
            8, order_by = District), tbelow20LAG8W = lag(tbelow20,
            8, order_by = District), tbelow15LAG8W = lag(tbelow15,
            8, order_by = District), pabove1LAG8W = lag(pabove1,
            8, order_by = District), pabove50LAG8W = lag(pabove50,
            8, order_by = District), pabove100LAG8W = lag(pabove100,
            8, order_by = District)) %>% drop_na()

# Data creation for mapping

poly1 = readShapeSpatial("Mozambique_admin1.shp", IDvar = "NAME1")
mapmal = mozmal
mapmal$Province = fct_collapse(mapmal$Province, Maputo = c("MAPUTO",
    "MAPUTO CIDADE"), `Cabo Delgado` = c("CABO DELGADO"), Gaza = c("GAZA"),
    Inhambane = c("INHAMBANE"), Manica = c("MANICA"), Nampula = c("NAMPULA"),
    Niassa = c("NIASSA"), Sofala = c("SOFALA"), Tete = c("TETE"),
    Zambezia = c("ZAMBEZIA"))

# Question 4 maps

excpt = as.data.frame(tapply(mapmal$malaria, list(mapmal$Province),
    sum))
exrainTot = as.data.frame(tapply(mapmal$rainTot, list(mapmal$Province),
    sum))
extavg = as.data.frame(tapply(mapmal$tavg, list(mapmal$Province),
    sum))


plotstats = as.data.frame(cbind(excpt, exrainTot, extavg))
colnames(plotstats) = c("Cases", "Rainfall", "Temperature")
mapdat = SpatialPolygonsDataFrame(poly1, plotstats)
```

```r
# Question 5 map

cpt = as.data.frame(tapply(mapmal$cpt, list(mapmal$Province,
    mapmal$Epiyear), sum))
colnames(cpt) = c("cpt10", "cpt11", "cpt12", "cpt13", "cpt14",
    "cpt15", "cpt16")
plotstats2 = as.data.frame(cbind(cpt))
breaks.qt = classIntervals(plotstats2$cpt16, n = 6, style = "quantile",
    intervalClosure = "right")

polydat = SpatialPolygonsDataFrame(poly1, plotstats2)

# Lags

# Avg Temp Two Week lag POISSON
pois <- lm(cpt ~ tavgLAG2W * Region, data = mal2W)
rss2.2 <- c(crossprod(pois$residuals))
mse2.2 <- rss2.2/length(pois$residuals)
rmse2.2 <- sqrt(mse2.2)

# Four Week lag POISSON
pois <- lm(cpt ~ tavgLAG4W * Region, data = mal4W)
rss4.2 <- c(crossprod(pois$residuals))
mse4.2 <- rss4.2/length(pois$residuals)
rmse4.2 <- sqrt(mse4.2)

# Eight Week lag POISSON
pois <- lm(cpt ~ tavgLAG8W * Region, data = mal8W)
rss8.2 <- c(crossprod(pois$residuals))
mse8.2 <- rss8.2/length(pois$residuals)
rmse8.2 <- sqrt(mse8.2)

# Total Rain Two Week lag
mal2W$sqrtRain <- sqrt(mal2W$rainTotLAG2W)
# POISSON
pois <- lm(cpt ~ sqrtRain * Region, data = mal2W)
rss2.2 <- c(crossprod(pois$residuals))
mse2.2 <- rss2.2/length(pois$residuals)
rmse2.2 <- sqrt(mse2.2)
# Four Week lag
mal4W$sqrtRain <- sqrt(mal4W$rainTotLAG4W)
# POISSON
pois <- lm(cpt ~ sqrtRain * Region, data = mal4W)
rss4.2 <- c(crossprod(lm2$residuals))
mse4.2 <- rss4.2/length(lm2$residuals)
rmse4.2 <- sqrt(mse4.2)
```

```r
# Eight Week lag
mal8W$sqrtRain <- sqrt(mal8W$rainTotLAG8W)
# POISSON
pois <- lm(cpt ~ sqrtRain * Region, data = mal8W)
rss8.2 <- c(crossprod(pois$residuals))
mse8.2 <- rss8.2/length(pois$residuals)
rmse8.2 <- sqrt(mse8.2)

# Temp above 30 2 weeks POISSON
pois <- lm(cpt ~ tavgLAG2W * Region, data = mal2W)
rss2.2 <- c(crossprod(pois$residuals))
mse2.2 <- rss2.2/length(pois$residuals)
rmse2.2 <- sqrt(mse2.2)

# 4 weeks POISSON
pois <- lm(cpt ~ tavgLAG4W * Region, data = mal4W)
rss4.2 <- c(crossprod(pois$residuals))
mse4.2 <- rss4.2/length(pois$residuals)
rmse4.2 <- sqrt(mse4.2)

# 8 weeks POISSON
pois <- lm(cpt ~ tavgLAG8W * Region, data = mal8W)
rss8.2 <- c(crossprod(pois$residuals))
mse8.2 <- rss8.2/length(pois$residuals)
rmse8.2 <- sqrt(mse8.2)

# temp above 35 2 week POISSON
pois <- lm(cpt ~ tabove35LAG2W * Region, data = mal2W)
rss2.2 <- c(crossprod(pois$residuals))
mse2.2 <- rss2.2/length(pois$residuals)
rmse2.2 <- sqrt(mse2.2)

# 4 week POISSON
pois <- lm(cpt ~ tabove35LAG4W * Region, data = mal4W)
rss4.2 <- c(crossprod(pois$residuals))
mse4.2 <- rss4.2/length(pois$residuals)
rmse4.2 <- sqrt(mse4.2)

# 8 week POISSON
pois <- lm(cpt ~ tabove35LAG8W * Region, data = mal8W)
rss8.2 <- c(crossprod(pois$residuals))
mse8.2 <- rss8.2/length(pois$residuals)
rmse8.2 <- sqrt(mse8.2)

# temp below 15 2 week POISSON
pois <- lm(cpt ~ tbelow15LAG2W * Region, data = mal2W)
```

```r
rss2.2 <- c(crossprod(pois$residuals))
mse2.2 <- rss2.2/length(pois$residuals)
rmse2.2 <- sqrt(mse2.2)

# 4 week POISSON
pois <- lm(cpt ~ tbelow15LAG4W * Region, data = mal4W)
rss4.2 <- c(crossprod(pois$residuals))
mse4.2 <- rss4.2/length(pois$residuals)
rmse4.2 <- sqrt(mse4.2)

# 8 week POISSON
pois <- lm(cpt ~ tbelow15LAG8W * Region, data = mal8W)
rss8.2 <- c(crossprod(pois$residuals))
mse8.2 <- rss8.2/length(pois$residuals)
rmse8.2 <- sqrt(mse8.2)

# temp below 20 2 weeks POISSON
lm2 <- lm(cpt ~ tbelow20LAG2W * Region, data = mal2W)
rss2.2 <- c(crossprod(lm2$residuals))
mse2.2 <- rss2.2/length(lm2$residuals)
rmse2.2 <- sqrt(mse2.2)


# 4 weeks POISSON
lm2 <- lm(cpt ~ tbelow20LAG4W * Region, data = mal4W)
rss4.2 <- c(crossprod(lm2$residuals))
mse4.2 <- rss4.2/length(lm2$residuals)
rmse4.2 <- sqrt(mse4.2)


# 8 weeks POISSON
lm2 <- lm(cpt ~ tbelow20LAG8W * Region, data = mal8W)
rss8.2 <- c(crossprod(lm2$residuals))
mse8.2 <- rss8.2/length(lm2$residuals)
rmse8.2 <- sqrt(mse8.2)
```