# Climate, Governance and Disease

### **Environmental and Institutional Determinants of Malaria Spread in Africa**

# # Introduction

With crucial implications on public health and society development, the spatial-temporal variation of **malaria infection and mortality rates** in central Africa have received growing attention. Previous studies investigated the correspondence of malaria disease with climate conditions such as annual precipitation and daily maximum temperature. However, more detailed environmental conditions such as vegetation density, dominant tree species and surface water area may affect the habitats and propagation of mosquitos, the major vector of infection of malaria. In addition to these natural environmental conditions, control measures of human society also play crucial roles in observed malaria rates. Such control measures can be closely related to the quality of governance and demographics. In this study, we propose to investigate (1) the spatial-temporal patterns of reported malaria infection and mortality rates, and (2) their relationships with the environmental and socio-political variables. Malaria rates will come from published dataset for counties in central Africa. Environmental variables are to be obtained from satellite remote sensing datasets and global land surface model outputs; and geo-coded socio-political variables will be gathered from multiple sources. Spatial Gaussian process models will be implemented, and informative variables will be identified using the Bayesian Information Criteria.

**# Background**

\*\*Climate and Vegetation\*\* Vegetated ecosystems are considered vulnerable under warming temperature and increasing frequency and intensity of extreme ecosystem disturbances such as droughts, storms, fires and pest outbreaks @settele2014terrestrial. These climatological changes and perturbations may gradually alter plant physiological properties including stomatal kinetics [@brodribb2009evolution,lammertsma2011global], physiological strategies [@scheiter2009impacts, hawkes2008soil], phenology [@garonna2016variability, buitenwerf2015three] and ecosystem productivity [@piao2008net, keenan2013increase, drake2016carbon], or lead to a regime shift after reaching the tipping point [@scheffer2009early], with the new state exhibiting different dynamics, sensitivities to environmental conditions and ecological services. The consequent changes in vegetated ecosystems have the potential to alter the terrestrial carbon sink strength, local and global climate [@settele2014terrestrial], hence influencing ecological functions and human society.

As noted in the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) [@settele2014terrestrial]: "Uncertainty in predicting the response of terrestrial and freshwater ecosystems to climate and other perturbations, particularly at the local scale, remains a major impediment to determining prudent levels of permissible change." Given the multidimensional nature of climate-vegetation feedbacks, studies of the combined influence of multiple simultaneous factors are needed. Historical data provide opportunities to investigate empirical relations between environmental conditions and vegetation dynamics.

\*\*The Politics of Climate Change\*\* The recent decade has seen the emergence of a literature on the politics of climate change. Researchers studies the topic from a variety of theoretical perspectives. Researchers on the politics of climate change policies, as reviewed in @Bernauer2013, attempt to understand four important research topics: (1) institutional designs that facilitate or hinders international cooperation of on climate issues; (2) domestic factors that shape climate-related policymaking; (3) the role of civil society in climate change politics; (4) socio-political consequences of climate change. In comparison, the community of scholars studying conflict and war attempts to understand how the climate and whether is related to explains and predicts war through different channels. As reviewed in @Brown1990, a large stream of literature studies how climate, affect the risk of both interpersonal and inter-group conflicts. Overall, the review article argues that contemporaneous temperature increases interpersonal conflict by 2.4 percent and intergroup conflict by 11.3 percent. However, the review cautions that evidence we know so far are mixed and scholars face challenges in empirical studies in due to issues with model specification. More generally, climate change shown to have social and economic impact. Recent \*Science\* and \*Nature\* articles use rich geo-coded data and existing research to show that temperature changes have impact on health, economic productivity (agriculture, energy, trade), demography (violence, migration, women's welfare) [@Carleton2016; @Burke2015].

Studies understanding the particular link between \*climate\* and \*conflict\* in \*Africa\* constitutes a a particularly well-received literature in recent years. As one of an earlier contribution to the literature, @Hendrix2012 and @Hsiang2013 finds that rainfall variability has statistically significant effect on instances of political conflict of various scales, as it can disrupt the economy and provoke social tension. @Loughlin2012 finds that extremely warm weather increases the risk of conflict in Eastern Africa, based on spatial-temporal data from 1990 to 2009. Beyond the regional trend, the global warming is evidently associated with the increased conflict risk, as a study finds the El Nino/ Southern Oscillation can explain 21 percent of the civil conflict since 1950 [@Hsiang2011].

**# Data**

In this section, we describe our data. We collect our data from a variety of data collection effort of environment, demographics and governance statistics. The first challenge we face is that the variables in their respective original datasets are aggregated into geographic units of different sizes. With the objective to balance computational tractability and the precision of measures, we clean and merge two datasets: (1) a country-level areal dataset where each country-year has one observation; (2) a grid-year dataset, the spatial resolution is 0.5 degree. In the table below, we show an overview of variables and their sources.

**## Malaria**

We obtain country-level measures of the severity of Malaria spread from 2010 to 2016. Below we plot the data of 2012 in the map. The maps suggest spatial autocorrelation. We will test it in a later section.

**## Data cleaning and imputation**

The environmental and socio-political variables have different spatial resolution and temporal coverages. The environmental variables with finer spatial-temporal resolutions were all resampled to a 0.25 degree resolution and a monthly interval. Data of civil conflict with specific coordinates were assigned to the 0.25 degree grid they belong to, which were then aggregated together to obtain the total number of conflicts within each grid. The gridded population data extends to 2012. For the years after 2012, population was imputed as being equal to that in 2012 with the assumption that population changes little within the latter 4 years. A few coastal and island pixels have missing climate variables, which could be because large fraction of sea surface in those pixels. These pixels were excluded from the dataset. The mosquito presence data was provided in @kyalo2017geo, a most recent and comprehensive dataset from meta-analysis on malaria vector in Africa. The dataset contains the year, location and species of mosquitoes reported in previous literatures. The total number of mosquito presence, regardless the species, was aggregated to each 0.25 degree grid.

**## Exploratory Analysis**

In this section, the number and location of mosquito presence within a year is modeled. We first examine if spatial correlation exists in the point-referenced data. As suggested by the following variogram for each year, mosquito presence might be spatially correlated within ranges of around 1 degree in years of 2011 and 2013. The association appears unclear for other years.

Moran's I was also computed for mosquito presence within each year, with the weight matrix consisting of the inverse of distance between each pair of points. Moran's I's range from 0.015-0.082 across the years, indicating a weak spatial association of mosquito presence. As such, normal generalized linear models are used without incorporating spatial dependence.

## Model and results

The point-referenced counts of mosquito presence are modeled using a Poisson regression, weighted by population.

$$

\log \left( E(Y|X)\right) = \alpha + \beta'X

$$

where $Y$ is the counts of mosquito presence weighted by population and $X$ include predictors. As noted in the conceptual graph, mosquito presence can be affected by both environmental conditions including surface water area, vegetation density, rainfall, air temperature and humidity, and sociological development extent that can be partly represented by population, distance to capital, number of conflicts and nightlights. These are included as candidate variables. All possible combinations of up to three of the candidate variables were used to fit the data from all years during 2010-2016. Then models with the lowest Bayesian information criterion are listed below.

The results show that annual mean air temperature (Tair\_mean) and its intra-annual variation (Tair\_sd) were selected in the four models with the lowest BIC. This is consistent with previous studies highlighting the crucial roles of air temperature and its seasonality on mosquito quantities in Africa [@zhou2004association; @pascual2006malaria]. In addition, air humidity and vegetation index also contributes to the estimation of mosquito presence The densest vegetation (NDVI\_90) rather than the average of vegetation density (NVDI\_50) within a pixel was also identified as an informative predictor, possibly due to the fact that dense vegetation provides more favorable habitat for mosquitoes. Notably, population is positively correlated with mosquito presence, which might be a result of more data collections at locations with larger population. Mosquito presence is also positively associated lower social stability indicated by larger number of deaths in civil conflicts.

As the values of BIC are close to each other, mosquito presence rate at locations without observation was predicted using the ten models with the lowest BIC via model averaging. That is, the predicted value by each of the ten models were averaged to obtain a surface of mosquito presence. This continuous estimation of mosquito presence rate within each year were then used as a variable to be used to estimate malaria cases and deaths within each country.

**# Areal Data Analysis**

In this section, we present results of the spatial patterns of malaria spread in Africa. As discussed in Section 3, data of malaria infection and deaths is only available at the country level from 2010 to 2016. We take the subset of 2012 statistics as this is the latest year where we available of our socio-political variables. This section is organized as follows. We first present results of our exploratory analysis where we test the spatial correlation using different versions of operationalization. In the second part, we choose an outcome, the number of malaria cases diagnosed, and fit a variety of models to predict it. We start with a set of linear models and then attempt hierarchical models to account for the nature of the outcome as a count variable.

**## Exploratory Analysis**

We explore the spatial correlations of four outcome variables of interest regarding the severity of Malaria in African countries: number of cases, number of cases per capita (cases / population), number of deaths, number of deaths per capita, and the mortality rate. As is shown in the table below, all outcome variable shows significant \*positive\* spatial autocorrelations. We set value of the weight matrix as binary (whether two countries share border). We use this simple measure instead of distance because countries are large and transportation in many area of Africa is not well advanced. So we assume the contagion of the diseases travel very far. In particular, the \*number of cases diagnosed\* has the strongest spatial correlation. Thus, we choose the number of cases as our main outcome variable of interest.

**## Analysis**

We fit a linear Conditional Autoregressive (CAR) model to predict the number of malaria cases. We take the logarithm of the outcome so that the data fits the a linear model. Formally,

\begin{gather\*}

\mathbf{y} \sim N(\mathbf{X}\beta, \Sigma\_{CAR}) \\

\Sigma\_{CAR} = \sigma^2(\mathbf{D} - \phi \mathbf{W})^{-1}

\end{gather\*}

\*\*No Predictors\*\* We start with a CAR model without predictors, which shows poor predictive power. Below we show a comparison of the actual distribution of malaria cases and the fitted values. As the figure shows, the prediction of the number of cases is off for many of the countries -- it only captures the high malaria risks in a two central African countries (the countries colored yellow and light blue). Thus, spatial autocorrelation is unable to explain the variation of malaria. We next fit models we the set of predictors we consider important, as introduced in previous sections.

\*\*Adding Predictors of Demographics, Governance and Environment\*\* We use a set of predictors to improve the model. After model selection, we present model with best performance (i.e. highest log-likelihood and lowest AIC). The model includes three variables: mean nighttime lights, number of observed mosquitoes and air humidity. We plot the size of the coefficients of the scaled variables below. The results show that nighttime light is negatively associated with the number of malaria cases. This is in line with theoretical prediction, as higher nighttime lumination is associated with better quality of governance, which indicates that it can be also capable of controlling diseases. On the other hand, air humidity is positively associated with the number of malaria cases, because higher humidity facilitate propagation of mosquitoes and provides more favorable habitats for mosquitoes. However, find the counterintuitive result that the number of mosquitoes observed has no correlation with malaria risk (which is also the case across all other models we experimented). We puzzle about reasons, which we will discuss in the conclusion.

The best model with governance and environmental predictors shows improved predictive power. As we show in the graph below, the fitted value better captures the variation of malaria cases.

**# Conclusion**

Large spatial variation exists in malaria cases and deaths in different African countries. For example, based on the data in 2012, countries with large number of malaria cases and deaths are mostly concentrated in the eastern Africa and several countries in the western Africa; the number of malaria deaths is the highest in central Africa. In this study we examined the spatial connection in point-referenced data of malaria related mosquito presence and areal data of malaria cases and deaths by county. We found that mosquito presence exhibits little spatial connection whereas malaria cases and deaths by country have clear spatial structures. Among all the considered environmental and sociological variables, air temperature is robustly identified to be positively related with the number of mosquito presence. Other variables including air humidity, vegetation density and population also contributes to spatial variation in mosquito presence. Across countries, denser vegetation, higher air humidity and larger rainfall amount are associated with higher malaria cases. Notably, higher values of night light is negatively associated with malaria cases, which indicates that higher quality of governance may help control the number of malaria cases.

One interesting issue arises from our analysis is that spatial correlation was almost not identified from the data of mosquito, which is the major vector of malaria transmission. However, malaria cases exhibited clear spatial association. This is contradictory to our initial hypothesis that neighboring countries may share low/high number of malaria cases due to transmission vectors of mosquitoes. One possible reason might could be the uncertainty in mosquito dataset. As the mosquito dataset comes from meta-analysis, the choice of locations for mosquito research may not be random, hence undermining the representativeness of the spatial structure of mosquito presence. Hence the mosquito presence data used here may not serve well as an intermediate variable connecting environmental conditions and malaria.

Further study may involve more detailed investigation on the influence of environmental and socio-political variables on malaria cases, deaths and mortality rate. Predictive model may be established and serve as a tool to predict future malaria using projected climate conditions, possible sociological changes and policies.