BAMM: a Bayesian Adaptive Mathematical framework for Malaria transmission dynamics

This paper introduces BAMM: a Bayesian Adaptive Mathematical framework for Malaria transmission dynamics. BAMM begins by utilising an existing malaria mathematical compartmental model (Yang [6]) & informs the parameterisation of the model in a Bayesian setting by deriving an indirect mapping between both social economic factors & medical data and the relevant transmission dynamics posterior parameter estimates. The framework provides a software tool in the form of a dashboard that allows researchers to posit & simulate any number of feasible scenarios. Making assessing either isolated events or the interaction effect of events trivial. The framework allows for hypothesis testing on a wide variety of topics including (but not limited to): temperature changes, access to treatment, changes in inequality measures & growth in GDP.

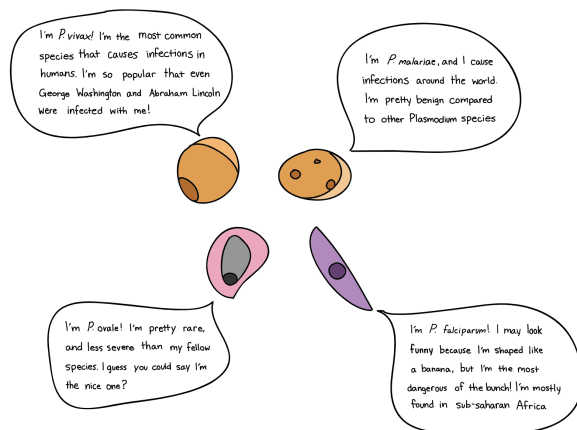University of Cape Town

Zach Wolpe   2020   WLPZAC001

Table of Context

Biology & Background

Malaria is an infectious disease caused by a parasite - Plasmodium - that invades red blood cells & liver cells, transferred to humans by the bite of an infected Anopheles mosquito.

Species of the plasmodium parasite

I'm *P. vivax*! I'm the most common species that causes infections in humans. I'm so popular that even George Washington and Abraham Lincoln were infected with me!

I'm *P. malariae*, and I cause infections around the world. I'm pretty benign compared to other Plasmodium species

I'm *P. ovale*! I'm pretty rare, and less severe than my fellow species. I guess you could say I'm the nice one?

I'm *P. falciparum*! I may look funny because I'm shaped like a banana, but I'm the most dangerous of the bunch! I'm mostly found in sub-saharan Africa.

There are 4 different species of the parasite, characterised by the figure besides this text.

Malaria life cycle

1. The parasite is injected into a human's blood stream by an infected mosquito in the form of *sporozoites* - which travel in your blood until they infect the liver cells.

2. Once in the liver, the parasite multiplies & matures over a period of 5-16 days (depending on the species) producing *merozoites* (the young adult form of the parasites). For some species *merozoites* can remain dormant for months before maturing.

3. After maturing the *merozoites* m enter the blood steam to infect red blood cells - where they begin asexual reproduction producing thousands of merozoites that are released over 1-3 days. This release causes fever like symptoms. Some merozoites are not asexual & do take on gender (gametocytes).

4. If another mosquito bites the infected person, it digests the gametocytes, which fuse together & burrow into the walls of the mosquito's stomach while they mature inside oocysts.

5. After a 8-15 day maturation period, the oocysts burst & *sporozoites* are released into the mosquito's body cavity. They then travel to the mosquito's saliva gland, primed for the next transmission.
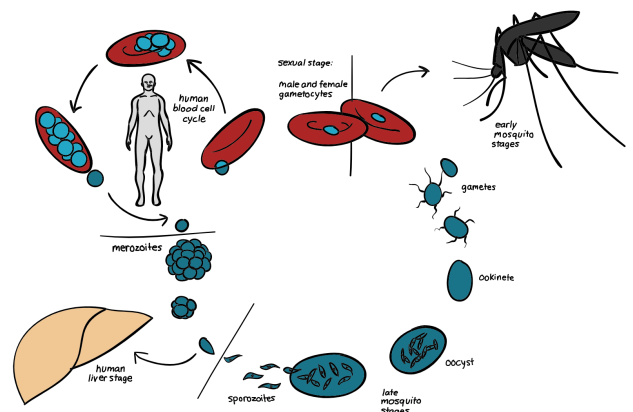
Symptoms

The disease has a 1-2 week onset before any symptoms arise - whilst the above biological process is undertaken. All species usually causes 'flu-like' symptoms (headache, fever, nausea and muscle pain) that directly coincide with increasing number of parasites in the body.

Malaria paroxysms, also known as febrile attacks, often follow a few days later. Typically lasting 4-8 hours & characterised by 3 stages:
- Cold stage: 15-to-60 minutes of shivering & feeling extremely cold.
- Hot stage: 2-6 hours, following the cold stage, of experiencing a high fever.
- Sweating stage: 2-to-4 hours where one's fever drops dramatically.

These paroxysm episodes ebb & flow as the disease reproduces & new merozoite parasites enter the bloodstream. This cyclical process occur roughly 48 hours for all species bar P. malariae - which has a 72 hours cycle.

Malaria tends to be a chronic disease - meaning without treatment symptoms will often relapse weeks or months after the initial onset, as dormant merozoites activate. These relapses may continue for decades.

Whilst P. vivax, P. malariae & P. ovale can cause severe illness, they are rarely fatal. P. falciparum is the most dangerous species - causing the majority of fatalities.

Severity classification

The severity of malaria cases is classified into two grounds:

Uncomplicated
- The individual could experience 'flu-like' symptoms, paroxysms & jaundice (yellowing skin complexion)
- These infections are still dangerous & uncomfortable however seldom fatal as they are effectively treated with antimalarial drugs.
- In the absence of treatment uncomplicated cases can be fatal.

Severe
- Additional to that which is experienced in uncomplicated cases, severe cases can be lethal as the individual can experience organ related issues. Repository complications leading to the inability to breathe normally; or circulatory problems leading to exceedingly low blood pressure; or severe anemia or brain issues resulting in fatigue or a decrease in consciousness/coma.
- Severe malaria is thus characterised by systemic symptoms.
- Individuals suffering from severe cases need hospitalisation & intensive care.

Diagnoses

protection Malaria can be diagnosed by inspecting a blood sample of an individual. A formula is added to a drop of blood & examined under a microscope, thereafter the various species can be identified by an expert.
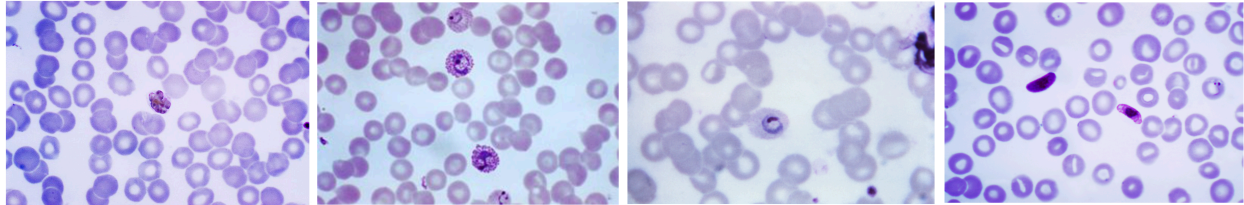
| P. malariae | P. vivax | P. ovale | P. falciparum |

Following a physical exam, a malaria diagnosis is done by undertaking lab tests. The two most methods involving drawing blood from an individual & thereafter:

1. Visually analysing a blood smear to assess/identify parasites (as shown above). This method is the simplest & safest however it requires a trained individual on hand.
2. RDT (Rapid Detection Test) is a test kit available to diagnose malaria. Analogous to a common pregnancy test, a dip-stick can be used to effectively diagnose malaria (& identify the species) without the need for human expertise (after drawing blood).

Prevention

Though no vaccine is available, one can take considerable efforts to prevent the onset of malaria. Medication is available to prepare one's body to fight off the disease if one will be travelling to a malaria inhabited location - unfortunately these pharmaceuticals do not work for permanent protection. Repellents & mosquito nets are still the most effective prevention available.

Treatments

Once diagnosed a combination of antimalarial drugs - artemisinin-based combination therapy (ACT) are prescribed. The ACT combination is crucial as incorrectly specified combinations could result the parasite mutating ???. Combinations are based on a variety of factors:
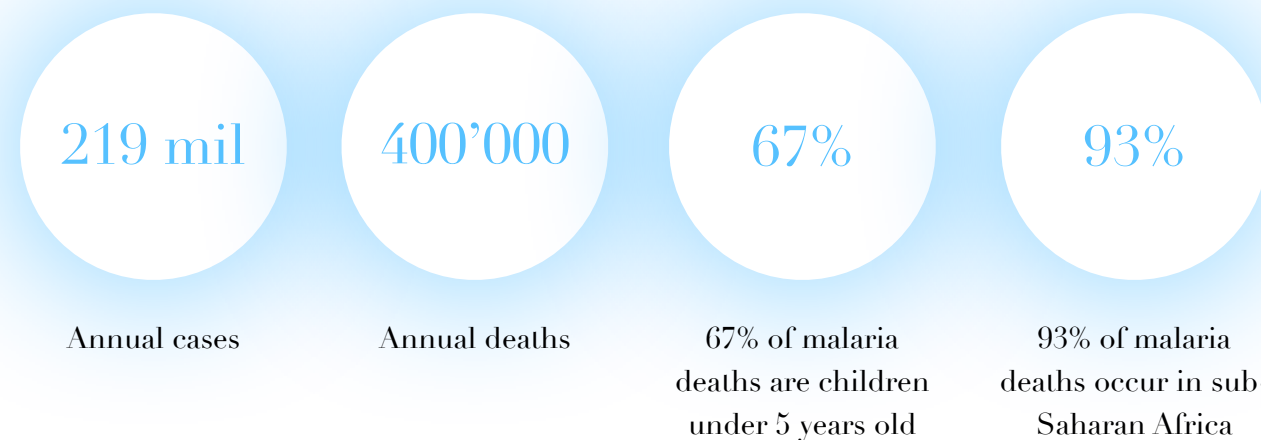
- Type of parasite
- Location of contraction
- Severity of symptoms
- Other illnesses, conditions or medications/allergies
- Whether or not the individual is pregnant
- Knowledge of existing parasite resistance to available antimalarial drugs

Early diagnosis & treatment is imperative as it can significantly reduce symptoms & prevent death.

Global Fight Against Malaria

In order to adequately model malaria & conduct scenario assessment, one ought to fully understand what has currently been attempted &/or achieved in the global fight against this malicious parasite. In 2016, The WHO committed to a 14 year plan called the *Global Technical Strategy* for Malaria that aims to eradicate malaria entirely in the next few decades, aiming for a 90% reduction in incidence (on a 2015 base) by 2030. A few years on & we are able to assess the current trajectory & viability of the plan. Whilst many targets have not been met, it it is still, albeit ambitious, an achievable goal.

Malaria is a severe health & economic burden on society. It is fundamentally a problem for societies most vulnerable people: the very young & poor. $3 billion is dedicated to the fight against malaria, as vulnerable populations grow the per capita spend is expected to decrease substantially. The subsequent economic impact is staggering, as so many are left unable to work due to the disease. The return on investment, both morally & financial, of successful malaria eradication is undeniable.

219 mil	400'000	67%	93%
Annual cases	Annual deaths	67% of malaria deaths are children under 5 years old	93% of malaria deaths occur in sub-Saharan Africa

Mega-trends effecting malaria

Malaria eradication is considered to be somewhat of a moving target - given the individual societal complexities that greatly influence the trajectory of the fight against the disease. WHO has identified a number of 'mega-trends' that substantially impact the malaria incidences, including changes in: climate, demographics, migration, population growth, urbanisation & land use and land coverage. More nuanced trends, of which the effects on malaria transmission dynamics are less clear, were also identified: woman empowerment, growth in access to information technology, education and political structure.

WHO eradication strategy

Whilst I will not provide great detail on the WHO's strategy to fighting malaria as it is superfluous to our needs, a succinct summary of the plan is available as diagram 1 in the appendix.
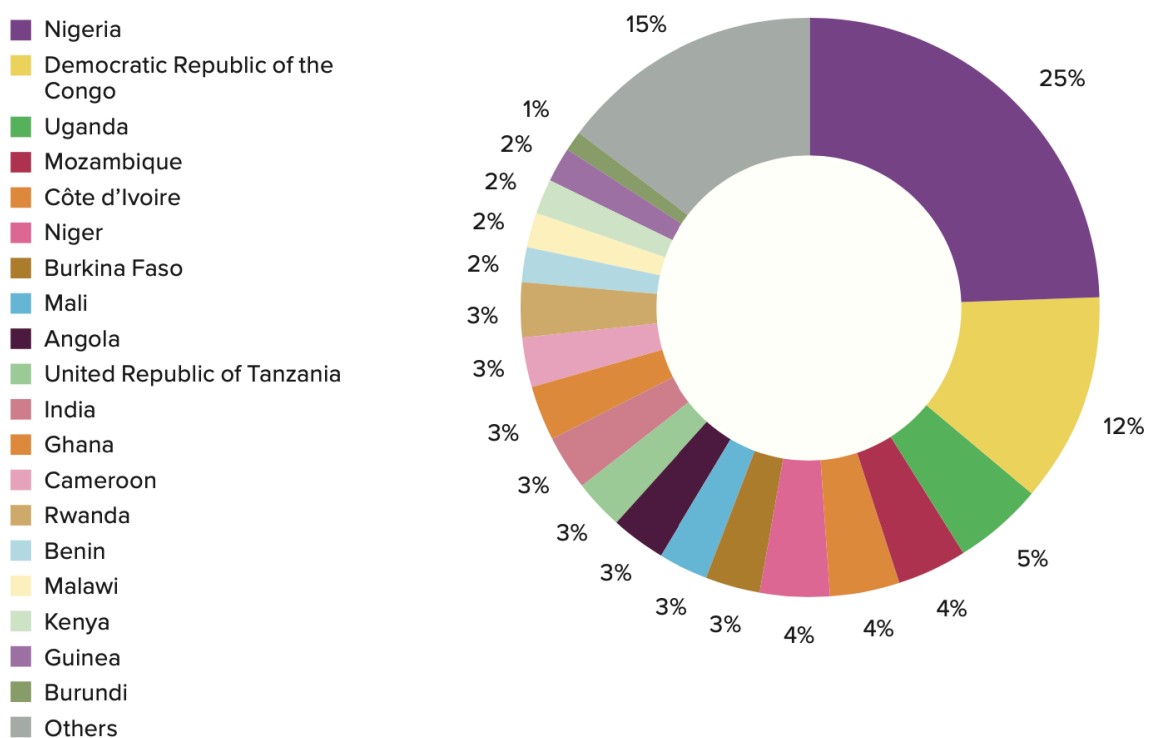
Country specific intervention

Malaria is far from a uniform problem, each infected country hosts a unique range of challenges, yielding a universal approach unjustified. All leading research on the disease emphasises the need for unique, adaptive solutions for each target region. As well as ecological, geographical & social-economic factors, the above-mentioned mega-trends account for these differences.

Further, as mentioned, this distribution of malaria's impact is far from uniform. A few, poor, countries bare the majority of the burden, as seen in figure 3 below [2].

As such, our analysis will focus on the the 3 most severely burdened countries: Nigeria, DRC & Uganda.

Fig. 23. Estimated country share of total malaria cases, 2018

Legend:
- Nigeria
- Democratic Republic of the Congo
- Uganda
- Mozambique
- Côte d'Ivoire
- Niger
- Burkina Faso
- Mali
- Angola
- United Republic of Tanzania
- India
- Ghana
- Cameroon
- Rwanda
- Benin
- Malawi
- Kenya
- Guinea
- Burundi
- Others

Research Objective: BAMM

Mathematical modelling

Mathematical frameworks have been used to model diseases for over a century [3]. If mathematicians are able to adequately decompose transmission dynamics - understanding their constituents - biologist & policy makers are better able to tackle disease elimination & eradication. Additional to providing a pragmatic understanding of a given disease, mathematical models are equally useful in their ability to extrapolate current circumstances to forecast potential future scenarios - allowing for more prudent preparation & decisive action. Mathematical models describe the relationships between components in a system - often used to illustrate co-dependencies, interactions & causal effects.

In specifying the model's structure, it is implicit that the modeller have some understanding of the relationship between variables of interest (parameters) - normally derived from the theoretical literature of domain expertise.

Statistical modelling

In resent years, a byproduct of the big-data revolution, statistical models have grown increasingly applicable. Dichotomous with mathematical models, statistical models attempt to learn probabilistic distributions (as apposed to static values) of parameters of interest based on data.

Hybrid models

Whilst both modelling approaches have been extensively implemented individually - a movement towards complex modelling systems that rely on both mathematical & statistical methods is developing [3].

A particular variant of these combinations is of interest in the diversity of malaria modelling across different regions. There is an excepted consensus among malaria experts that systems/regions/populations are heterogeneous & ought to be modelled in the light of their unique circumstances [2]. The availability of large, accurate, datasets may allow for realistic, probabilistic, parameterisation of mathematical models, thus leading to the objective of this research, to develop:

BAMM: a Bayesian Adaptive Mathematical framework for Malaria transmission dynamics.

BAMM is *Bayesian* in that it relies on Bayesian statistics for parameter estimation; *Adaptive* in that domain specific datasets can be used to learn model parameter distributions & a *Mathematical framework for Malaria transmission dynamics* in that the model being parameterised is a compartmental malaria model.
Essentially using Bayesian statistics to parameterise a malaria transmission compartmental model.

Why BAMM?

model The goal of this approach is to provide a modelling framework that can be adapted to any particular region, 'customisable' so long as the region as access to good data. This framework was chosen for a myriad of reasons:

1. Each population facing malaria is riddled with unique circumstances & should be considered accordingly. Domain specific datasets allow for domain specific modelling.

2. Data availability allows for more accurate model specification. We should rely on data as much as possible, over-and-above domain expertise, as counterintuitive results may be found. Expertise should be used to reason data, not data to reason expertise.

3. Whilst burdensome to specify adequately, a Bayesian framework is advantageous for copious reasons:

   1. Confidence bounds can be learnt effectively

   2. In the light of new available or updated data, the model can be reparameterised by using the previous *posterior* distribution as a *prior*, the new data as the *likelihood* and thus allowing for an updated *posterior* that incorporates all findings.

   3. Bayesian estimations is also advantageous as in the event that parameters are reasonably known by theoretics, theoretical parameters can be used as priors in the Bayesian posterior derivation - effectively incorporating both domain expertise & relationships learnt in the data.

Research objectives

The goal of this paper is 3 fold:

1. *Specify & describe a flexible mathematical framework for malaria modelling.*

2. *Derive a Bayesian posterior for parameters of interest.*

3. *Implement a case study & assessing the solution.*

Mathematical Model Derivation

A baseline mathematical model needs be derived. Mathematicians have been modelling malaria transmission dynamics for over a 100 years, & whilst advancements in modern medicine, computational equipment & ecological findings have improved these models, the original fundamental structure remains [3]. Originating in the Ross model in 1911, the general structure imposed is model two populations independently - humans & mosquitos - & derive their interaction as the parasite completes it's life cycle.

Models generally trade-off computational feasibility & realism. In our case wish to utilise a sufficiently complicated model to allow for unique region parameterisation - yet maintaining computational efficiency & interpretability. The Bayesian parameter posterior estimation is a further computational burden, the aggregate computation should be considered.

Our model architecture should adequately consider the factors that distinguish countries. WHO identifies a number of direct & indirect influences on a countries malaria eradication efforts: population growth, urbanisation, climate change, land use & land cover changes (LULCC), social economic status & migration trends [4]. Indirect factors are equable notable - woman's empowerment, access to technology, political structure, etc - however they yield complex, interactive effects that are difficult to formulate either theoretically or mathematically.

A further region specific parameter is development of malaria immunity among exposed populations. Communities repeatedly exposed malaria (as in many rural African communities) immunity can be gradually, naturally, established [5]. The nature of this immunity calls for immunity classification as a spectrum & not merely a binary response [5].

We are going focus one three distinctive features of a region & implement a relevant model to capture these features - however the framework is readily extended to more encompassing frameworks:
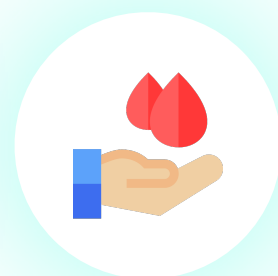1. Developed immunity
2. Climate change (temperature increase)
3. Social economic status (access to health & other resources)

Immunity   Global Warming   Social Econ Status

Yang Model

Hyun Yang proposed a model that captures the spectrum of immunity & derives temperature dependent parameters [5]. This model is then extended to to include a climate change & social economic status [6]. I will implement these models as soundly address the problem at hand. They are parameterised by theoretical research - allowing for a simple Bayesian extension for parameterisation, thus we arrive at *BAMM*. Notation from the original paper is used [5].

The model is parameterised by 2 distinct populations (that of humans & mosquitos). The model can be visualised in the below schema:
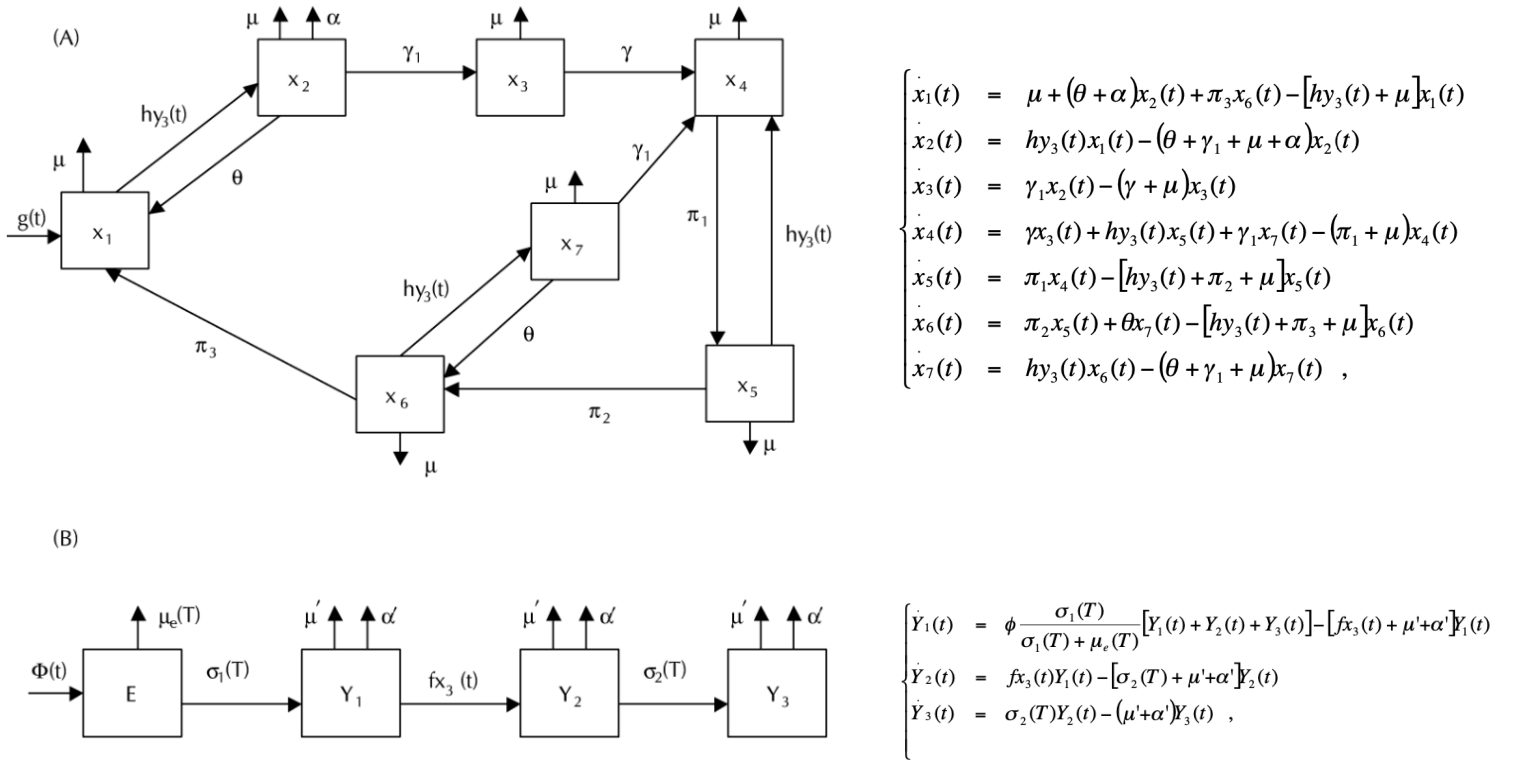

$$
\begin{cases}
\dot{x}_1(t) &= \mu + (\theta + \alpha)x_2(t) + \pi_3 x_6(t) - [hy_3(t) + \mu]x_1(t) \\
\dot{x}_2(t) &= hy_3(t)x_1(t) - (\theta + \gamma_1 + \mu + \alpha)x_2(t) \\
\dot{x}_3(t) &= \gamma_1 x_2(t) - (\gamma + \mu)x_3(t) \\
\dot{x}_4(t) &= \gamma x_3(t) + hy_3(t)x_5(t) + \gamma_1 x_7(t) - (\pi_1 + \mu)x_4(t) \\
\dot{x}_5(t) &= \pi_1 x_4(t) - [hy_3(t) + \pi_2 + \mu]x_5(t) \\
\dot{x}_6(t) &= \pi_2 x_5(t) + \theta x_7(t) - [hy_3(t) + \pi_3 + \mu]x_6(t) \\
\dot{x}_7(t) &= hy_3(t)x_6(t) - (\theta + \gamma_1 + \mu)x_7(t) \quad,
\end{cases}
$$

$$
\begin{cases}
\dot{Y}_1(t) &= \phi \frac{\sigma_1(T)}{\sigma_1(T) + \mu_e(T)}[Y_1(t) + Y_2(t) + Y_3(t)] - [fx_3(t) + \mu' + \alpha']Y_1(t) \\
\dot{Y}_2(t) &= fx_3(t)Y_1(t) - [\sigma_2(T) + \mu' + \alpha']Y_2(t) \\
\dot{Y}_3(t) &= \sigma_2(T)Y_2(t) - (\mu' + \alpha')Y_3(t) \quad,
\end{cases}
$$

Figure - The schematic diagrams of the overall malarial transmission: human host (A) and mosquito vector (B).

Incubating/exposed (x_2) are those individuals who are infected but do not yet have gametocytes (cannot infected other mosquitos) - x_3 have circulating mature gametocytes. $x_4 \rightarrow x_6$ provide discrete placeholder for the spectrum of immunity [5]. The method of acquiring immunity is simplistic, ignoring the age of first infection & immunity development duration.

The paper derived two possible steady-state equilibrium solutions (when transition dynamics no longer change with time) [5]. The two solutions result in either a cyclical endemic or the disease dying off. The derivations are straightforward, though omitted as they are superfluous to our needs. Stability analysis - guaranteeing numerical stability of equilibria - is also provided in the paper.

11

Parameter Interpretation

Human hosts:

$x_1 = susceptible$
$x_2 = incubating/exposed$
$x_3 = infectious$
$x_4 = immune$
$x_5 = partially\ immune$
$x_6 = non-immune\ but\ with\ immunologic\ memory$
$x_7 = incubating\ after\ reinfection$

$\mu = natural\ mortality\ rate$
$\alpha = disease-induced\ mortality\ rate$
$\theta = natural\ resistence\ against\ malaria$
$\gamma_1 = average\ periods\ to\ initiate\ the\ production\ of\ gametocytes$
$\gamma = average\ periods\ to\ build\ up\ an\ effective\ immune\ response$
$\pi_1 = Rates\ at\ which\ protective\ immunity\ wanes$
$\pi_2 = rates\ at\ which\ partial\ immunity\ wanes$
$\pi_3 = rates\ at\ which\ immunologic\ memory\ wanes$

Notable:

$\frac{\sigma_1(T)}{\sigma_1(T) + \mu_1(T)} = probability\ of\ an\ egg\ transfirmation\ during\ \sigma_1(T)$

$g(t)\ and\ \Phi(t)$ are model names & bear no consequence.

Mosquito vectors:

$Y_1 = susceptible$
$Y_2 = incubating/exposed$
$Y_3 = infectious$

$\mu' = natural\ mortality\ rate$
$\alpha' = induced\ (insecticides\ etc)\ mortality\ rate$
$\phi = rate\ of\ oviposition$
$T = temperature$
$\mu_e(T) = rate\ of\ eggs\ becoming\ non-viable$
$\sigma_1(T) = cycle\ duration\ from\ eggs\ to\ adult$
$\sigma_2(T) = duration\ of\ sporogony\ (gametocytes\ to\ infective\ sporozoite$
$f = transmission\ rate$

Host-Vector Interaction Terms

$f = transmission\ rate$
$h = inoculation\ rate$

Basic Reproduction Ratio: R0

The model leads to an Basic Reproduction Ratio R0 formulation:

Which has four terms [5]:
1. Probability the individual will survive the latent period & be in the infected state.
2. Related to the number of susceptible mosquitoes infected with gametocytes by an infectious individual during his/her entire infective period.
3. Probability that the mosquito will survive during latent period.
4. The number of susceptible individuals infected with sporozoites by an infectious mosquito during it's entire infectious life.

Note that the first two terms related to the infection of susceptible mosquitos whilst the last two related to that of individuals [5]. Further, note that the immunity levels do not directly effect R0, however do so indirectly by delaying the recurrence an individual to becoming susceptible.

Model Parameterisation

Original paper results

The originating paper goes on to conduct a steady-state scenario analysis of the disease under varying social-economic conditions; temperature changes (due to climate change) & perceived level of risk in the community. Whilst I'll omit the full description for brevity - here are the key findings - the full discussion is available in the referenced paper [6].

Inputs were discretised into three categories for temperature (low, medium, high) & three categories for social-economic standing (low, medium, high) - resulting in 9 combinations. Three scenarios were then explored: describing a population as being either low, intermediate or high risk of malaria. In each scenario we wish to consider the distribution of compartments within both human & mosquito populations as well as the reproduction rate R0 to assess whether the scenarios steady state is either *disease eradication* or *endemic*.

Importantly, social economic factors have a far larger effect on the severity & likelihood of an endemic (described by R0) than temperature change. These factors manifest in good data collection (leading to timely decision making) & prompt access to testing & treatment for infected individuals.

The variation in R0, & thus importance, of these factors is far larger among high risk than lower risk communities.

The following parameter ranges were used in the original paper:

Table 1 - The values found in the literature for the parameters of the model. The symbols d and y stand, for days and years, respectively.

Parameter	Range	Mean
θ^{-1} (d)	$1 - 4$ [6]	2.5
γ_1^{-1} (d)	$15 - 19$ [5]	17
γ^{-1} (d)	$50 - 150$ [14]	100
π_1^{-1} (d)	$40 - 60$ [14]	50
π_2^{-1} (y)	$0.2 - 5$ [4]	2.6
π_3^{-1} (y)	$1 - 20$ [4]	10.5
μ^{-1} (y)	$50 - 55$ [4]	52.5
α^{-1} (y)	$2,450 - 2,964$	2,707
ϕ (eggs/d)	$25 - 65$ [5]	45
μ'^{-1} (d)	$10 - 14$ [5]	12
α'^{-1} (d)	$98 - 191.8$	144.9
$\sigma_1^{-1}(T)$ (d)	$10\ (31\ °C) - 26\ (20\ °C)$ [8,10]	18
$\sigma_2^{-1}(T)$ (d)	$8\ (31\ °C) - 22\ (20\ °C)$ [8,10]	15
$\mu_e^{-1}(T)$ (d)	$0.020\ (31\ °C) - 0.052\ (20\ °C)$	0.036

Note: The index numbers in the second column (range) refer to the bibliographic references listed in the end of the article.

$$\frac{\sigma_2(T)}{\sigma_2(T) + \mu'+\alpha'} \times \frac{h}{\mu'+\alpha'}$$

As specified by Yang:

$$\mu_e(T) = \sigma_1(T)\left(\frac{\phi}{\mu'+\alpha'} - 1\right)$$

Key Interaction parameters

$h = inoculation\ rate$
$f = transmission\ rate$

Parameter breakdown

Parameters were sourced from existing literature & are considered feasible epidemiologically sound. To better understand the model parameterisation, consider the following groupings. Temperature dependent parameters: $\{\mu_e(T), \sigma_1(T), \sigma_2(T)\}$; social-economic (& thus medical access) dependent parameters: $\{\theta, \gamma_1, \gamma, \pi_1, \pi_2, \pi_3, \phi\}$.

The interaction effects between mosquito (vector) & human populations are captured in $h \; and \; f$. Measuring the frequency & likelihood of individuals being exposed to mosquito populations, both $h \; and \; f$ are dependent on temperature (vectorial capacity); social-economic condition (bednets & deforestation) & climate [6]. These parameters are hard to estimate & are taken from the literature. They are used to categories a population into low, medium or high risk.

f is proportional to the number of infectious mosquitoes - equivalent to sporozoite rate measured in field surveys [6]. Below are the f and h values used in the paper analysis. We will adopt these parameter values as they need be estimated by experts. *Infection per day* are the appropriate units. We have widened the scope of these interaction parameters to better capture a dynamic range. The range extends f to 2 & h to 3. In replication the simulation the original tiny values of transmission & inoculation rates result in eradication in most instances - which is far from plausible.

Risk category	f: transmission rate	h: inoculation rate
Low	0.13	0.07
Intermediate	0.17	0.25
High	0.25	0.90

Before progressing to the Bayesian parameter estimation, we are going to fit this model.

Paper Replication & Discussion

This section sets the foundational work by replicating & discussing the paper's findings. The paper focuses on steady state (endemic or eradication) of environments over a range of contingencies. Key: SES: Social-Economic-Status; Temp: temperature.

Condition 1: Low risk, low temperature, strong economy & high prevalence

The first condition examined begins with extremely high prevalence (host & vector infections at over 90%) but is low risk community, with good SES infrastructure & low temperatures. Despite the initial prevalence in both host's & vectors, the disease is totally eradicated within a few years.

Condition 2: Medium Risk

The first condition examined exhibits rough average results across all
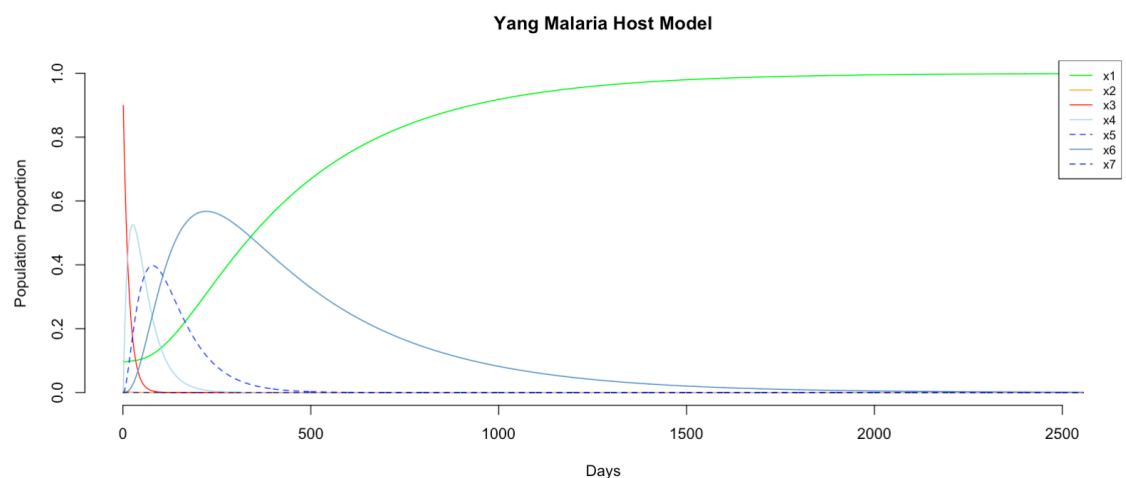
metrics. The resulting steady state is endemic - a fixed proportion of individuals & vectors being infected with malaria at any given time. This replication leverages average parameters for most estimates, however with an above average inoculation rate. Note the number of susceptible & infected are fairly constant, however there this great fluctuation in the range of immunological memory among individuals - perhaps indicative of the disease circulating rapidly. It may also be a mathematical consequence of unstable estimates - & warrant model specification examination.
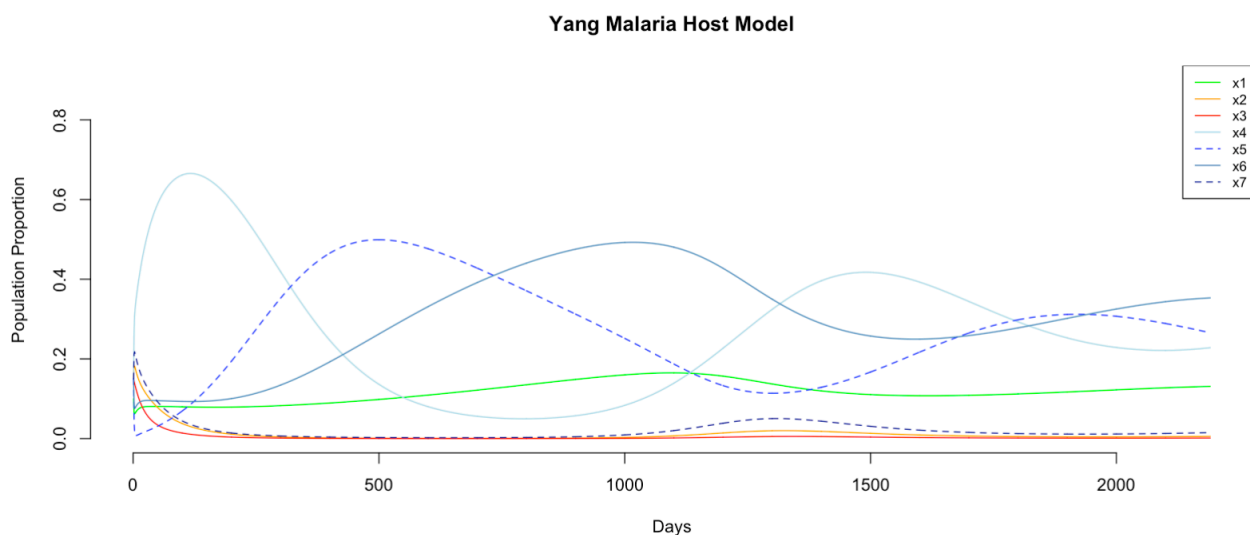
Condition 3: High Risk

In the final example, temperature is high, infection rates are high, interaction is high & SES is low. Evidently an even more serious endemic presents itself. The only reason the x1: susceptible class still persists is due to the high birth/death rate specification. One ought to consider the complexities of real life when examining these models, as in such an instance population would probably fall.

Yang Malaria Host Model

	x1
	x2
	x3
	x4
	x5
	x6
	x7

Par...

	Low ...
Risk	f=
SES	
Temp	Low
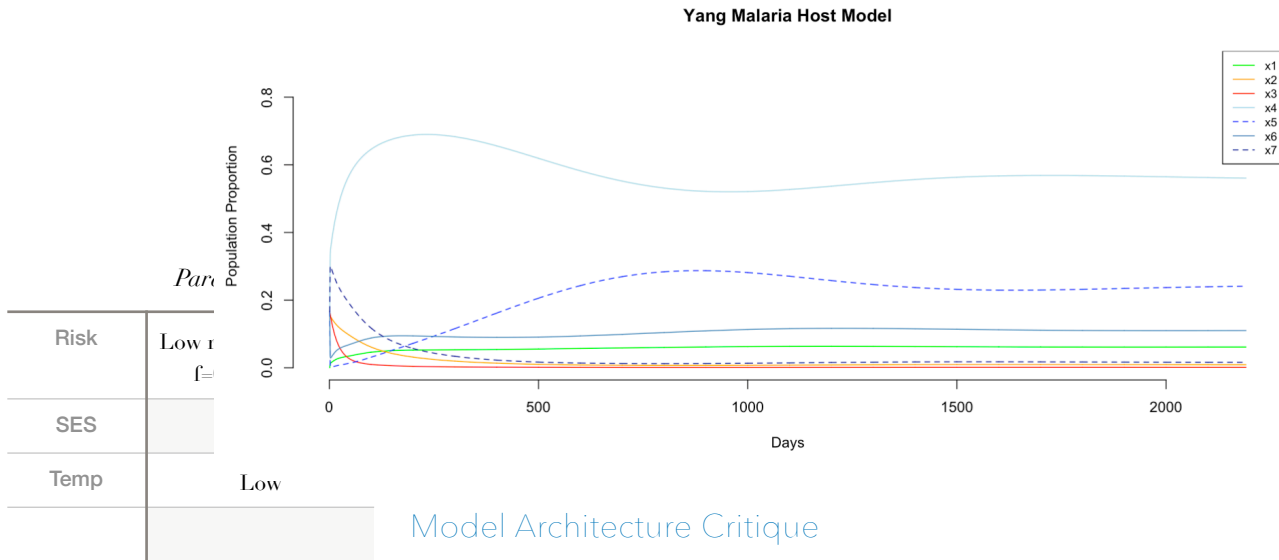
Model Architecture Critique

The mosquito population model is simplistic but realistic. Mosquito's are born susceptible & once infected carry the parasite until death. One ought not complicated the model further than necessary.

The human host model, however appears to possess some clear over-simplifications. Implicit in the model architecture is the notion of immunity being a linear scale - which is what the literature appears to suggest as immunity is built up gradually through severe exposure - however the lack of links between certain subpopulation nodes may be brought into question. x_1 (susceptible) $> x_2$ (incubating/exposed) $> x_3$ (infectious), thereafter the spectrum of immunity is examined $x_4 > x_5 > x_6 > x_7$ (immune $>$ partially immune $>$ non-immune with immunological memory $>$ incubating after reinfection). The sequence & movement between parameters is perfectly sounds, bar the inability of a candidate to move from x_7 (incubating after reinfection) to infection. One ought to consider adding a term to capture this movement (which is currently done purely indirectly).

Death rates are examined in an unconventional setting. Instead of having a 'recovered/deceased' class, individuals simple die of either natural μ or differential α (malarial infection) causes & are then 'added' back to the non-immune, susceptible class. Essentially holding population constant, implicitly holding birth & death rates equal. Although a perfectly natural assumption, the omission of a 'recovered/deceased' class mean one should be wary of examining the model outcome, as analysis is desensitised to death rates & it is not abundantly clear what proportion of individuals pass away from the disease. This isn't problematic in itself as it's superfluous to the model's purpose: to examine the effects of social-economic conditions, communal immunity & climate on R0, however it does warrant mention.

16

Bayesian Parameterisation: Temperature

Whilst it can appear tedious, the Bayesian approach is ideal in this setting as in reality incorporating prior knowledge is certainly advantageous. We may have domain expertise, previously collected data or regional specific information that can better inform our analysis.

Modelling Temperature

Temperature can be incorporated via two methods: average regional temperature used to parameterise the model or conducting a cosine or Fourier transform to add a temperature-temporal element to the transition dynamics of the model (capturing seasonality). We will conduct the former & leave the latter for later work. Unfortunately most historical weather datasets are restricted access on paid API's - as a make shift we will generate data that mimics our primary country of focus: Nigeria. This is a suitable alternative to illustrate the principle of Bayesian fitting. We will also generate a prior.
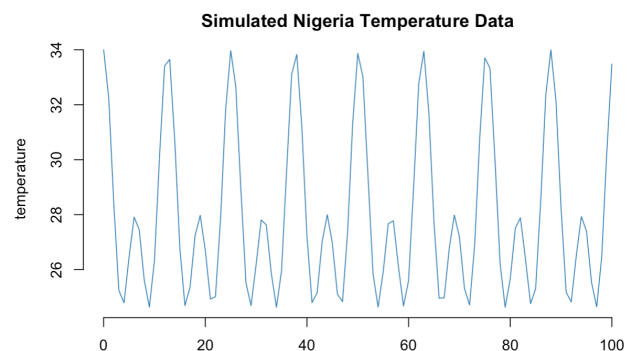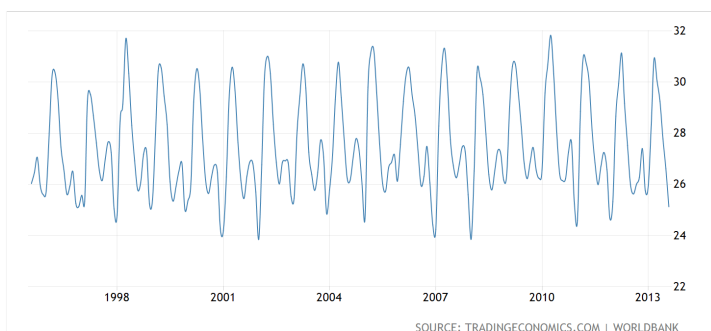
Prior

Let's assume the mean temperature over a region (West Africa) follows a Gaussian however we don't have much information. We want a mean temperature estimate, & the true data is bimodal (a consequence of seasonality). The min & max temperature's in West Africa are given as $18°C$ to $40°C$ - we formate a flat bimodal-gaussian prior over this range with a very high variance - essentially non-informative. The natural of the model allows one to tweak the prior in light of better information - the dashboard allows for these alterations.

Likelihood

The likelihood describes the observed data. Since we cannot obtain actual data without using a paid API, we will generate synthetic data to mimic the temperature patterns of Nigeria. The graph on the left is a snapshot of actual Nigeria temperature data, the graph on the right is our simulated data, drawn from the function:

$$f(x) = cos(\frac{x}{2})^3 + cos(x)^3 + 28$$

Simulated Nigeria Temperature Data

17

Posterior Derivation

As in any Bayesian model, we need to derive the posterior as a function of the likelihood, prior & marginal likelihood. In our model, we wish to find the posterior of *Temperature* : $\pi(\tau|x,\mu,\sigma^2)$ proportional to our prior (known assumptions of the regional temperature) & likelihood (Nigeria specific data).

$$posterior \propto prior \times likelihood$$

Below are two examples of fitting the Bayesian posterior to the data (likelihood) & the prior (domain expertise) for two different prior specifications.

The mean temperature values will be used to compute the temperature depended parameters in the differential equation model. The posterior can also be used to compute credibility intervals (confidence intervals in a Bayesian setting). Note the very low confidence in the posterior, resulting from the widely dispersed, multimodal prior & likelihood; as well as the incongruencies between the prior & likelihood.
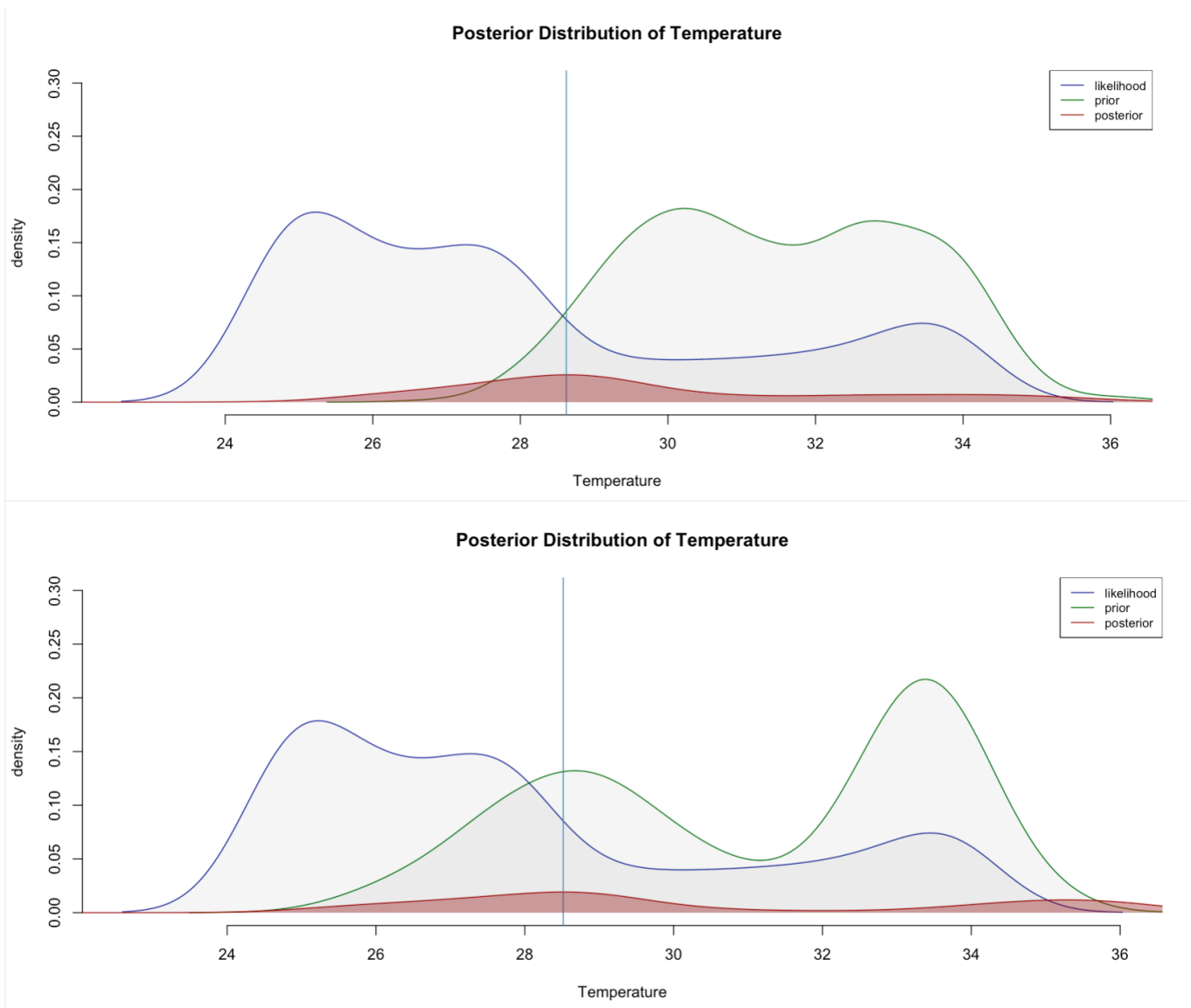
Bayesian Parameterisation: Social-Economic Status

Theoretical framework

We wish to fit social economic status as a predictor in our model, one caveat of the current model architecture is that direct, actionable, observable, social-economic-indicators are not given. Instead, Yang describes the related parameters as being *'roughly and indirectly'* related to social-economic status (SES) [6]. The relevant parameters are:

$$\{\theta, \gamma_1, \gamma, \pi_1, \pi_2, \pi_3, \phi\}$$

Parameters, whilst all influenced by biological & environmental factors, can be further broken down into two sections: parasital biology & human immunological response.

Parasital biology

 ϕ : oviposition of parasitical eggs

 γ_1 : average period to initiate the production of gametocytes

Human immunological response

 θ : a natural resistance rate against malaria

 γ : average period to build up an effective immune response

 π_1 : rate a which protective immunity is lost

 π_2 : rate a which partial immunity is lost

 π_3 : rate a which immunological memory is lost

Parasital parameters are, in fact, indirectly related to SES, however are a far more nuanced fashion. We will not extend the parameterisation of these variables, however we provide a Bayesian fitting approach, where the likelihood can be chosen by a domain expert & the prior is uniform over the range given by the literature.

Human immunological factors, however, are directly related to two core components: malaria exposure rates (as communities build up immunity overtime) & access to medical care. The former can be captured in the interaction terms in the model & can be ignored as they are accounted for. The latter needs be captured in the data - though two issues arise: access to precise medical data is often unavailable & social economic infrastructure also indirectly dictates access to medical attention. One can easily postulate how SES could influence decisions & access to medical care: for example, more developed societies are not only more educated & aware of these issues, their richer status allows them to focus on medical attention - whereas the deeply impoverished may only seek medical care in dire circumstances as they cannot afford to jeopardise work.

Bayesian parasital parameterisation

Parasitical related variable posteriors $\pi(\phi|X)$ & $\pi(\gamma_1|X)$ are derived with uninformative uniform priors & user specified (mean range default) likelihoods. This was only including to illustrate the at which the parameter estimable range could be updated in light of new information. The uniform prior has no influence on the posterior other than reducing confidence in the likelihood & limiting the permissible range. This feature is omitted from the dashboard as it serves no purpose.

Posterior Distribution γ1
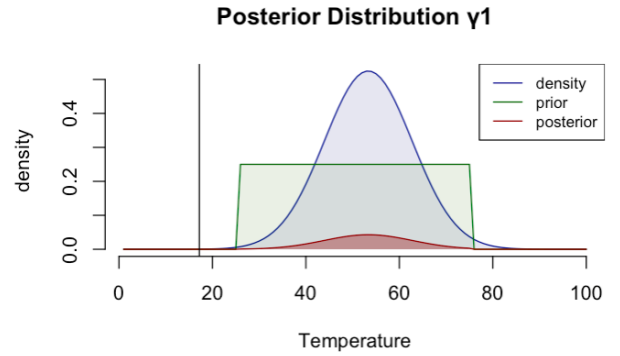
Human immunological response

Addressing our aforementioned issues of specific-data-scarify & the indirect interaction between SES & medical attention (beyond simply availability) **we can parameterise our SES immunological factor through their relationship with a varied spectrum of SES indicators.** Though possibly theoretical flawed, this unique & flexible approach is another example of the benefits of utilising the Bayesian framework - allowing for updating in-light of new information.

Given a set of correlated, multivariate, social economic indicators, we wish to parameterise:

$$\{\theta, \gamma, \pi_1, \pi_2, \pi_3\}$$

Direct estimation of these parameter is infeasible without very specialised medical records. Nonetheless, they all related indirectly to SES & thus are a function of SES. Theory dictates that these variables are also related to malaria incidence in a country. Both directly (higher incidence results in higher exposer to infectious mosquitos & medical treatment) & indirectly,

We will then, learn a relationship between malaria incidence & a representative collective of SES factors, & use this model as a pseudo-likelihood to inform the parameters of interest.

Data

I have curated a dataset from the World Bank with a number of countries social economic indicators as well as all malaria related figures. Though spanning over a wide range of countries, I visualised the data for the 3 primary countries of interest (Nigeria, DRC & Uganda). It's evident that a lot of the data is incomplete & possibly inaccurate, here are some important graphs. For brevity the remaining datasets are visualised in the appendix.
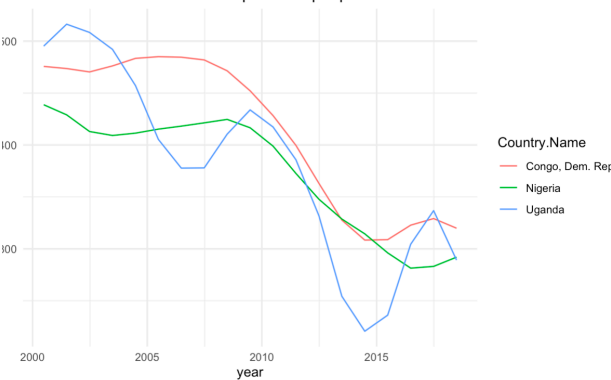
Malaria

% children under 5 that use insecticide treated bednets
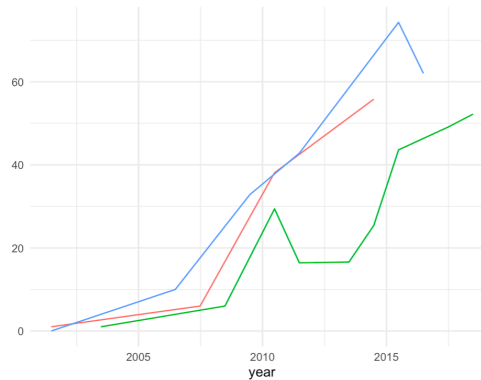
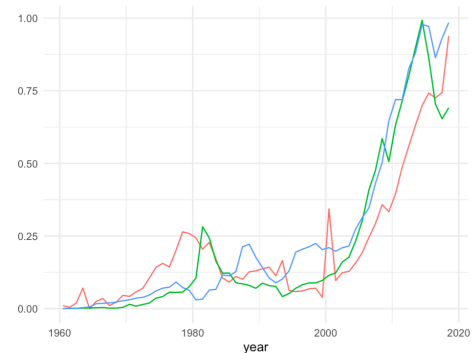% children under 5 with fever receiving antimalarial drugs
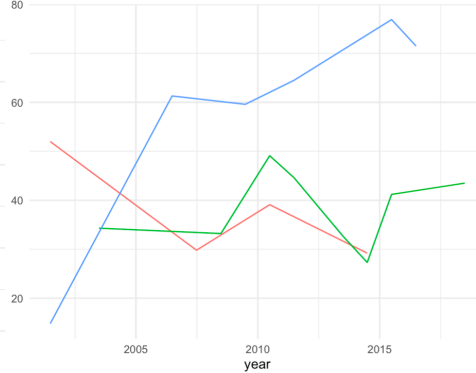
Incidence of malaria per 1000 people at risk

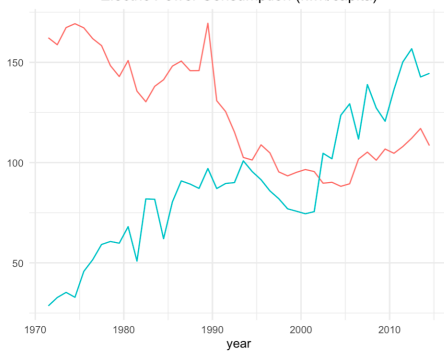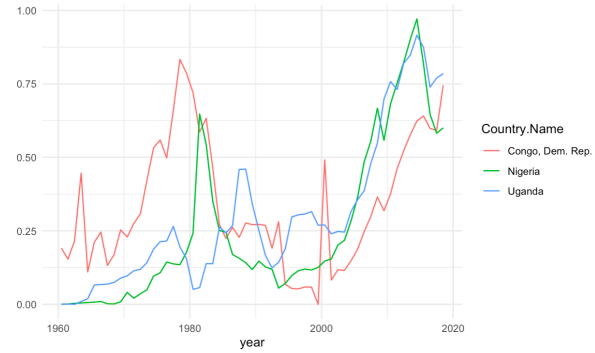### Economic Growth

Standardized GDP

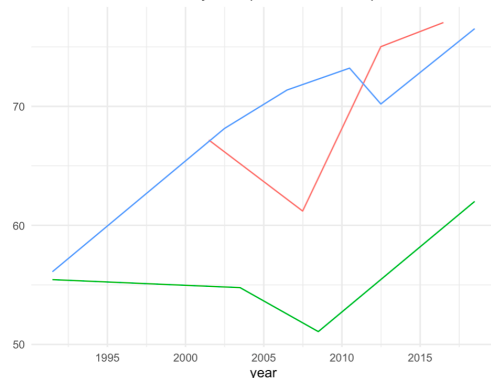Electric Power Consumption (kwh/capita)

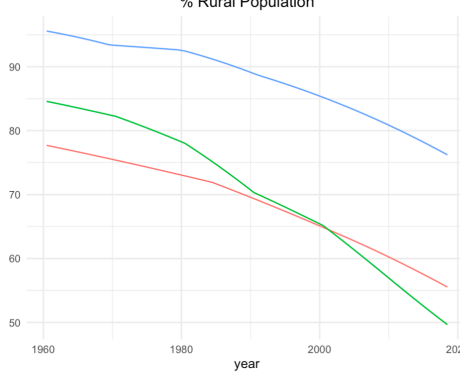Standardized GDP per Capita

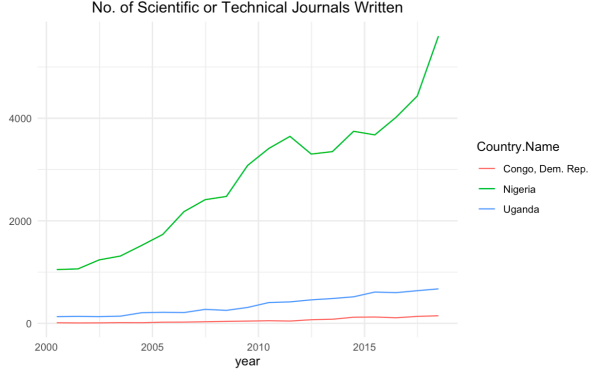### Education & Equality

Literacy rate (adults above 15)

% Rural Population

No. of Scientific or Technical Journals Written

Statistical Model

The mathematical model is centred around varying levels of immunity, we wish to use the data to better inform immunological memory transition dynamics. The only variable directly related to this is % children under 5 receiving antimalarial drugs.

The build up of immunological memory can be attributed too:
- Constant exposure to malaria (which is regional specific & highly dangerous, purely a consequence of negligence)
- Access to malarial treatments

We with to learn a posterior that best captures domain expertise in the prior & actual data in the likelihood to produce a feasible range of posterior distributions.

We thus fit a Bayesian posterior to each parameter.

Priors

Priors for each of the variables are considered to be unit variance gaussians with means to be specified for the country of interest. Allowing for domain knowledge representation over the parameters: $\{\theta, \gamma, \pi_1, \pi_2, \pi_3\}$.

Likelihood functions

Given the current data availability we cannot distinguish between factors that influence the varying degrees of immunological memory & resistance, as such a single pseudo-likelihood is computed as a proxy to update the individual priors. A standard regression likelihood assuming unbiased errors are normally distributed: $e \sim \mathcal{N}(0, \tau)$ is formulated.

Model Specification

Taking the available SES data, we formulate a dataset & design a model to capture the relationship between malaria incidence & SES factors. The following data is used:

y : incidence per 1000 people at risk

X_1 : literacy rates

X_2 : GDP growth

X_3 : % of population under 5 sleeping under ITNs

X_4 : % of population under 5 receiving anti-malaria drugs

X_5 : Rural population %

The data has a temporal element which needs be accounted for, as such, the response variable is modeled as a hierarchical/mixed effects model that is allowed random deviation across time & individual countries. This hierarchical model structure is similar to a standard generalized linear model, however different in that the covariance dependence structure of the covariates (individual countries) are allowed to vary overtime [7]. This mixed effect model is fit via REML (residual maximum likelihood, where the likelihood is given by [7]:

$$-2\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}_{\text{obs}}) = \log \frac{|\boldsymbol{L_\theta}|^2}{|\boldsymbol{W}|} + n \log(2\pi\sigma^2) + \frac{r^2(\boldsymbol{\theta})}{\sigma^2} + \frac{\left\| \boldsymbol{R}_X(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}) \right\|^2}{\sigma^2}.$$

Where θ is a unique covariance structure. The model fits the data very well. It is a primary a predictive tool as the complex covariance structure makes inference difficult, though doable. Before visualising the fit, we discuss one final imperative feature.

Scenario forecasts

An additional benefit of fitting the model around social economic factors is that it allows one to investigate the influence of changing those factors on malaria incidence in a given region. To this end, I developed an analytical tool that allows one to specific an increase or decrease in each of the social economic factors & the model is then used to predict how this would have impacted malaria incidence over a selected timeframe.

The parameters can be tuned from 0-5, which multiplied by the initial input parameter level & the used to forecast malaria incidence in that case. For example, changing the literacy rate to 0.8 results in moving the literacy rate to 80% of the original level. Similarly changing it to 1.2 movies it to 120% of it's original level. Adjustable as seen on the graph below.

This allows us to visualise the impact of a particular factor on malaria incidence.
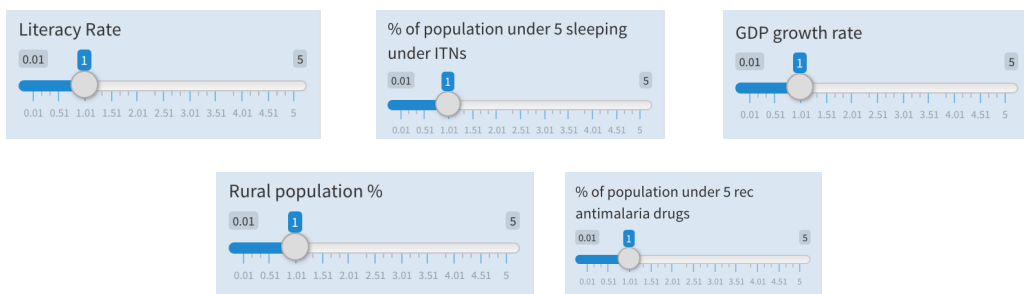
Model visualisation

Now that we have introduced the flexible forecasting tool, we can assess a given country over any timeframe over any parameterisation & see how the model would predict how a change in certain variable would effect malaria incidence. The the graphs below, blue depicts the prediction on the actual data, red the true value (trend in malaria incidence) & pink the model forecast for a given change in parameter (literacy rate etc).
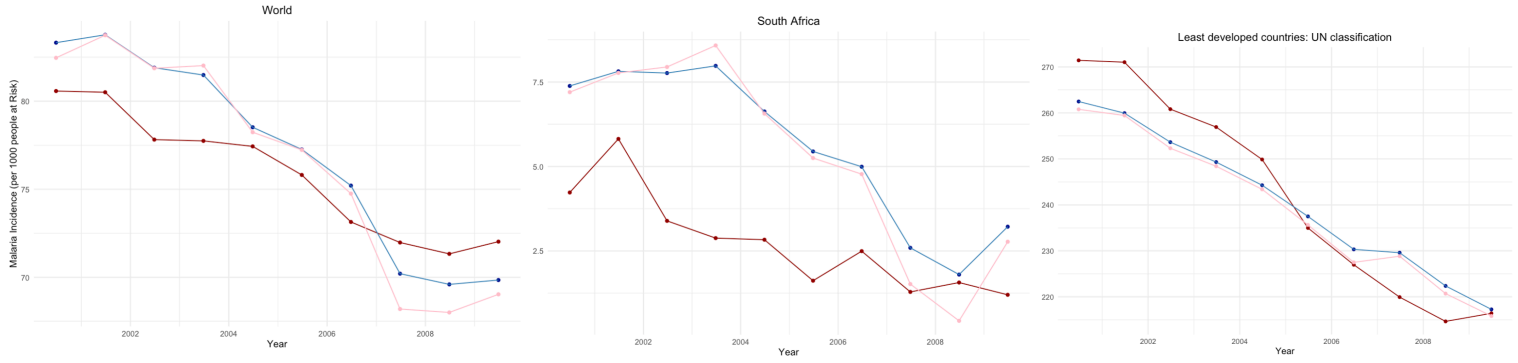
Since the model is trained on the entire dataset, severe deviation from global trend results in a poor fit. To negate this one could easy train the model for a specific region.

Inform immunological parameters $\{\theta, \gamma, \pi_1, \pi_2, \pi_3\}$

Now that we have a functional model that captures how SES factors influence malaria incidence, we want to used these variable to inform the parameterisation of the original mathematical model. This needs to be done indirectly so we need to define a mapping between the hierarchical model and the Yang (mathematical) model.

Recall the prior is set by a domain expert, as such we leave priors as the mean of each parameter range (though can be changed readily) & we want to use this hierarchical model as a pseudo-likelihood (a term I am using as it is not a real likelihood but we use it as the likelihood in the Bayesian sense).

Secondly, a key feature in the model is to assess how changes in the mixed effects model inputs change the Yang model. To achieve this we mad the change in forecast (if the mixed effects model is adjusted) to the mean of the parameters $\{\theta, \gamma, \pi_1, \pi_2, \pi_3\}$. I will further explain through an example:

Suppose we are looking at malaria incidence in sub-saharan African, we can the data to inform the yang model.

First we fit the model & received our generic prediction, we then wish to assess the impact of more a higher proportion of children receiving anti-malaria drugs & sleeping under ITNs (insecticide treated nets), we thus change these parameters to 2.5 each in our mixed effects model - resulting in a scenario test where 2.5 times the number of children use ITNs & have access to antimalarial drugs). The figure besides shows the effect of this change (difference between the pink & blue lines.
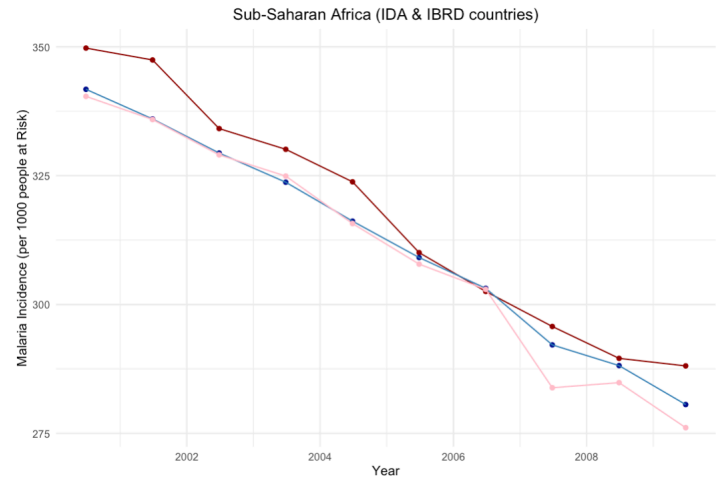

This slight improvement should be reflecting in the Yang model to assess transition dynamics. The difference between the two function (pink & blue) is then mapped to the posterior of each $\{\theta, \gamma, \pi_1, \pi_2, \pi_3\}$ variable. This is done by a transforming the different between the two to an equivalent change in mean for each of the $\{\theta, \gamma, \pi_1, \pi_2, \pi_3\}$ variable. Assuming a selected prior, this change in mean is mapped via a sigmoid function (selected to over-emphasise small changes & exhibit decreasing marginal return as one could expect in real life). This is done for all variable however it is visualised for theta.
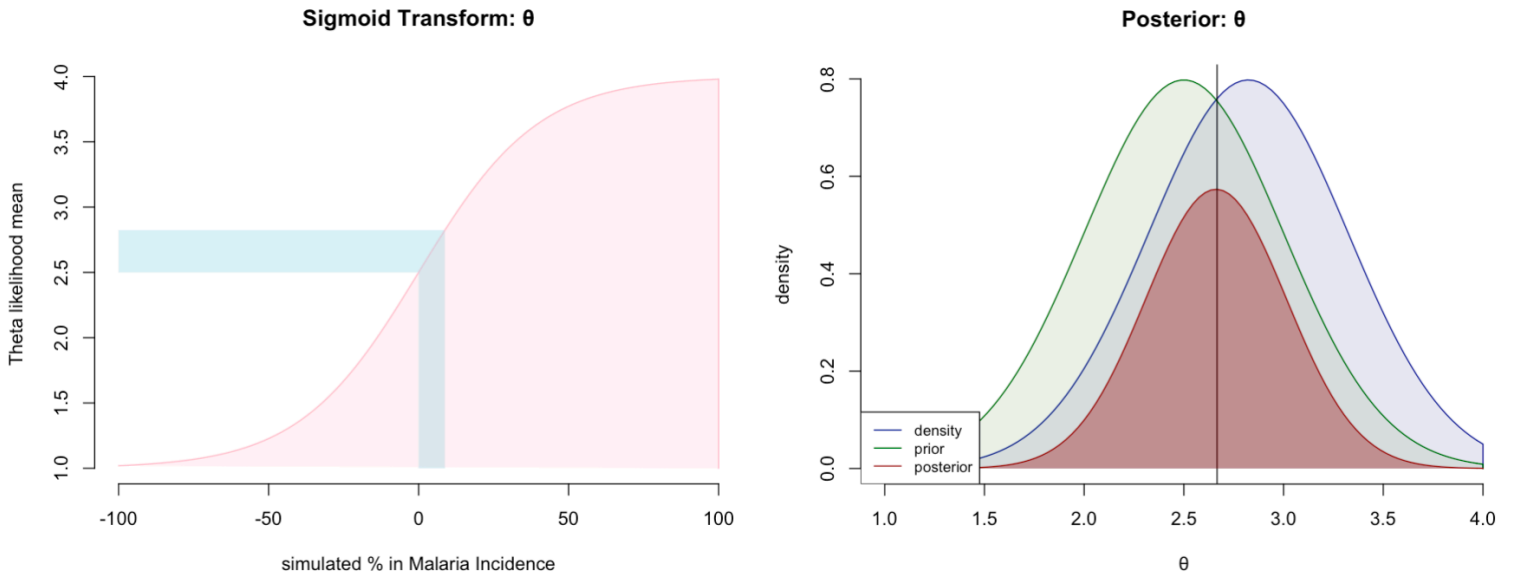
This mechanism allows us to use the change in fit as the pseudo-likelihood function to inform the posterior of each parameter $\{\theta, \gamma, \pi_1, \pi_2, \pi_3\}$. The posteriors are then used in the Yang model to predict the original transition dynamics between immunological groups.

25

BAMM: Scenario Analysis

As the purpose of this framework is to allow for domain expertise to be incorporated in an analysis of any region, a dashboard was created to allow for full flexibility. To illustrate how this can be done, we will analyse one scenario. The dashboard is available at.

Scenario:

Suppose we wish to fit a general model that using the trends in malaria incidence over the entire world to inspect immunological groupings amongst infected populations. We run the model for an 18 year period & assume that 85% population of interest have some exposure (immunological groupings) to malaria.

We further assume this to be an extremely high risk community with transition rate f=0.63 & inoculation rate h=0.81. All other parameters are set to mean levels, assuming now available expert knowledge.

We can now fit the model & assess the endemic state, notice how different immunological groups experience peaks & troughs, useful information for official making decisions as to when best to implement eradication schemes.

Yang Malaria Host Model
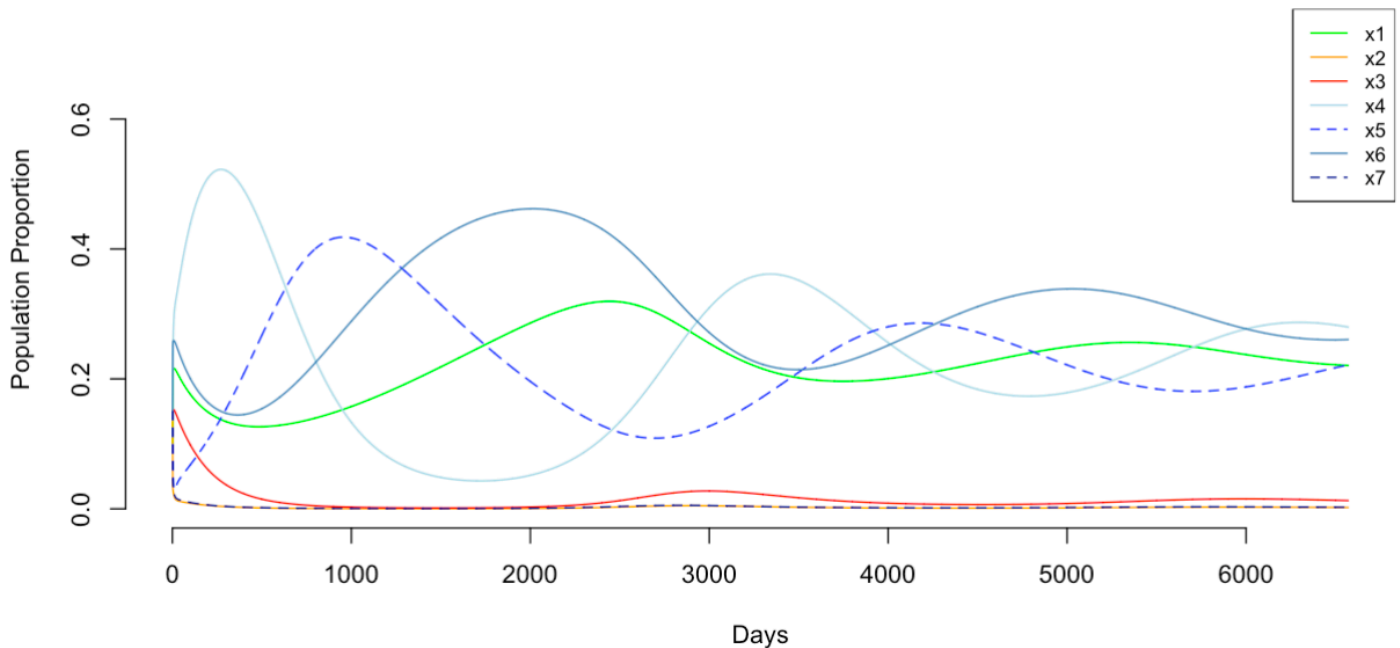
Now that we have a model, we wish to test an intervention scheme:

How would the transition dynamics differ if a lower proportion of individuals lived in rural settings?

To address our research question one can simply tune the parameter in the dashboard to observe the results. Note: dashboard available at www.zachwolpe.com/research. One simple adjusts the 'Rural population %' parameter to 0.9 to assess the outcome reducing the % of people living in rural settings by 10%. We then examine the dashboard:
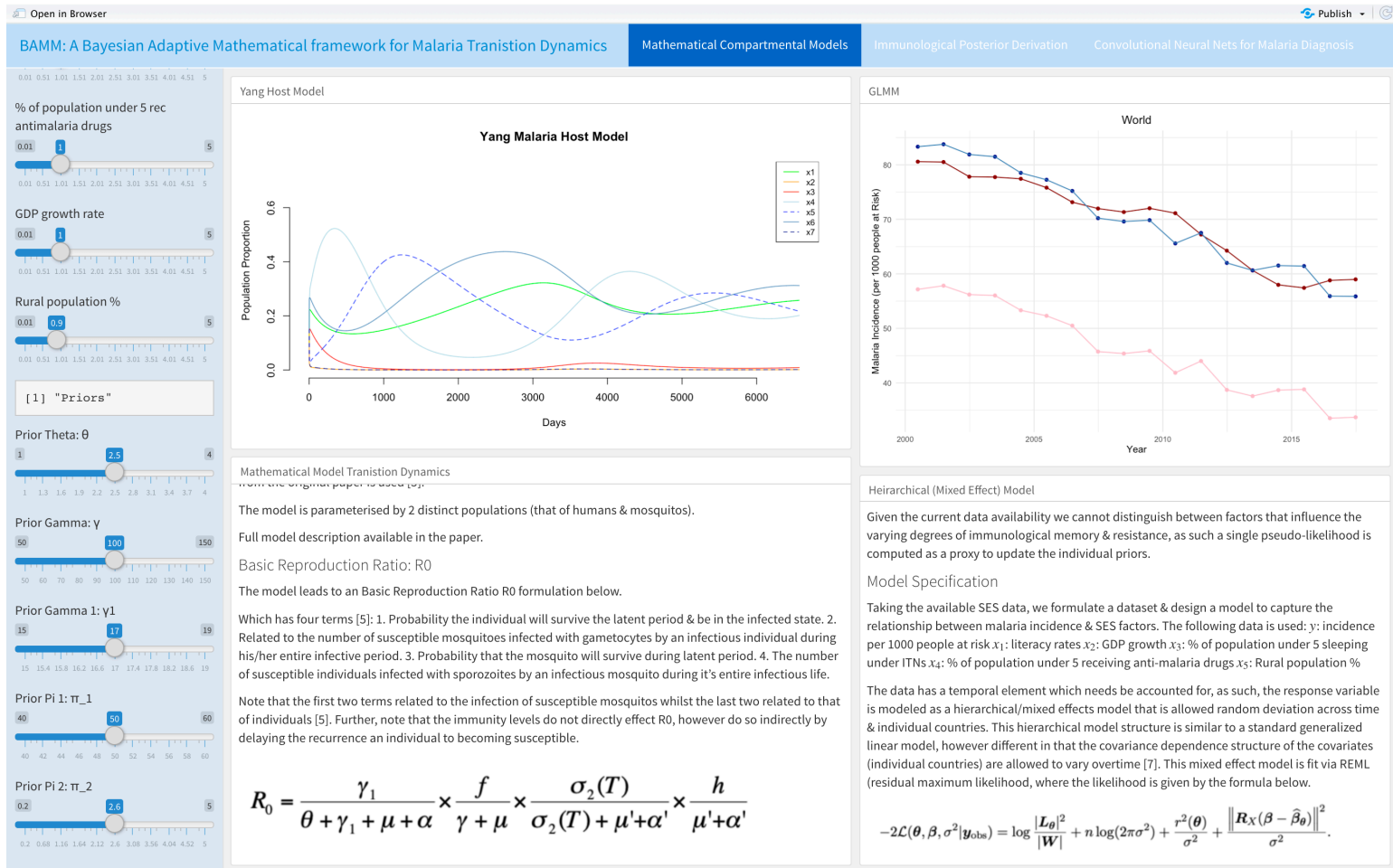
A chose this example for a reason, rural setting is an indication of inequality as many people living in rural domains are impoverished. Malaria is fundamentally a disease of the poor, as access to good treatment makes the world of difference.

The model predicts a major change in malaria incidence as a consequence of the fall in rural settlements. This then translates to updating the posterior of parameters involved in varying levels of immunological memory via the pseudo-likelihood, inspect the second page of the dashboard, provided below.

BAMM: A Bayesian Adaptive Mathematical framework for Malaria Tranistion Dynamics   Mathematical Compartmental Models   **Immunological Posterior Derivation**   Convolutional Neural Nets for Malaria Diagnosis

Sigmoid Transformation: Theta

Sigmoid Transform: θ

(plot: Theta likelihood mean vs simulated % in Malaria Incidence)

Teperature Generating Function

Simulated Nigeria Temperature Data

(plot: temperature vs time)

Theta Posterior

Posterior: θ

(plot: density vs θ — legend: density, prior, posterior)

Temperture Posterior

Posterior Distribution of Temperature

(plot: density vs Temperature — legend: likelihood, prior, posterior)

Theoretical background

Temperature

Whilst it can appear tedious, the Bayesian approach is ideal in this setting as in reality incorporating prior knowledge is certainly advantageous. We may have domain expertise, previously collected data or regional specific information that can better inform our analysis.

Modelling Temperature

Temperature can be incorporated via two methods: average regional temperature used to parameterise the model or conducting a cosine or Fourier transform to add a temperature-temporal element to the transition dynamics of the model (capturing seasonality). We will conduct the former & leave the latter for later work. Unfortunately most historical weather datasets are restricted access on paid API's - as a make shift we will generate data that mimics our primary country of focus: Nigeria. This is a suitable alternative to illustrate the principle of Bayesian fitting. We will also generate a prior.

Prior

Let's assume the mean temperature over a region (West Africa) follows a Gaussian however we don't have much information. We want a mean temperature estimate, & the true data is bimodal (a consequence of seasonality). The min & max temperature's in West Africa are given as °C to °C - we formate a flat bimodal-gaussian prior over this range with a very high variance - essentially non-informative. The natural of the model allows one to tweak the prior in light of better information - the dashboard allows for these alterations.

Likelihood

The likelihood describes the observed data. Since we cannot obtain actual data without using a paid API, we will generate synthetic data to mimic the temperature patterns of Nigeria. The graph on the left is a snapshot of actual Nigeria temperature data, the graph on the right is our simulated data, drawn from the function:

$$f(x) = cos(\frac{x}{2})^3 + cos(x)^3 + 28$$

Posterior Derivation

As in any Bayesian model, we need to derive the posterior as a function of the likelihood, prior & marginal likelihood. In our model, we wish to find the posterior of $\pi(\tau|X)$ proportional to our prior (known assumptions of the regional temperature) & likelihood (Nigeria specific data).

$$posterior \propto likelihood \times prior$$

Again we see the large transformation of the posterior of theta. Transition dynamics have been severely, greatly changing the scenario & providing using information policymakers & medical professionals.

The purpose of this example is simply to illustrate the vast flexibility of the BAMM system. Any number & variations of hypothesis could be tested.

Conclusion

BAMM: Bayesian Adaptive Mathematical framework for Malaria transmission dynamics allows one to simulate any variation of hypothesis to model the transition dynamics between immunologically distinct groups facing malaria. BAMM allows for domain expertise to be balanced with actual observable data to predict population transition dynamics between hosts & vectors carrying the malaria parasite.

As an addition, please see the section below, also available on the dashboard.

Convolutional Neural Nets for Malaria Diagnosis

Machine & statistical learning techniques can aid the fight against malaria in a multitude of ways - that extend far beyond parameterising mathematical compartmental models. To illustrate an example of other possible applications of statistical learning, consider the problem of accurate, complete diagnosis.

Malaria diagnosis

As aforementioned, in attempts to analyse any disease, we good, accurate data is imperative. Unfortunately it is almost always problematic. In the case of malaria, testing individuals in deeply impoverished communities is arduous. Assuming we can get access to members of those communities, our two testing options are:

- Utilise a testkit
- Blood examination by an expert

The former is the simplest & most reliable option, however resources are scarce & it may be infeasible to provide enough testing kits. The latter is only applicable if we have access to a physician.

Here I pose an alternative that elevates these resource constraints.

A Convolutional Neural Network

CNNs (convolutional neural networks) are a machine learning model architecture that using convolutional techniques to process high-resolution image data to allow for pattern discover. One can consider a colour image a (3 x n x p) dimensional sparse numerical matrix where n & p are the number of pixels & 3 represents the corresponding RGB values.

CNNs process images for pattern discover, often doing so far superior to human experts (one can think of a human expert as simpler running a more complex variant of this model architecture in their mind - we call this expertise). Here's a typical CNN architecture:
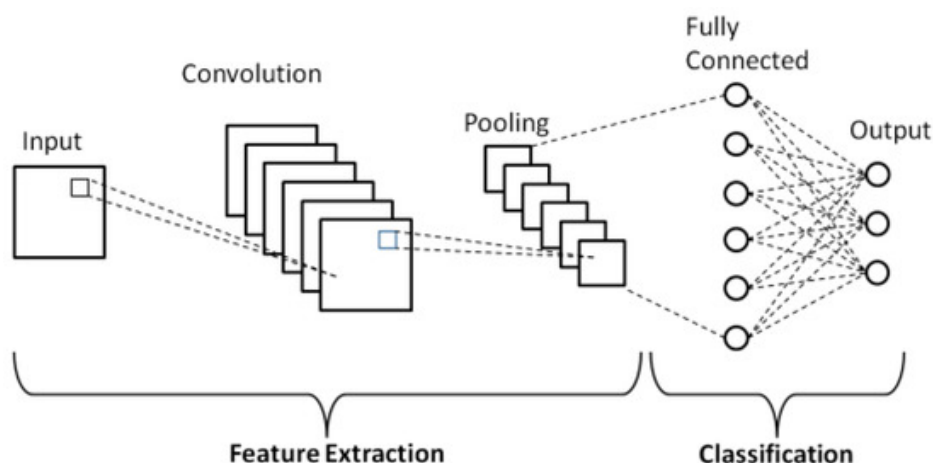
CNNs for Malaria Diagnosis

In the instance of malaria diagnosis, an expert would examine a blood sample & try to identify the parasite in the blood sample. Here we apply a CNN to detect the parasite, given an image.

Dataset

The labeled dataset utilised contains 27'558 images of blood samples - even split between infected & uninfected - and is available on the National Library of Medicine's (NLM) website [7].
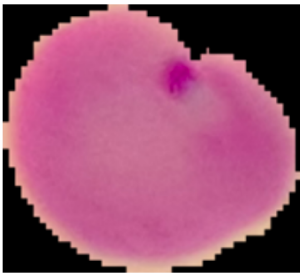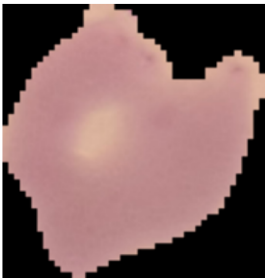
Results

The data was split into training & validation sets, thereafter the model was trained to detect parasites in the blood samples. The model architecture is available in the appendix & all code is available on the GitHub.

$$Accuracy = 90\%$$

90% accuracy was achieved - which may cause alarm of overfitting, thought the data was split into training-testing sets. I do not feel this result to be particularly alarming as the nature of the data is archetypical of the type of issues that CNN's perform particularly well on. Additional metrics were not examined as false-positive/negatives are roughly evenly distributed, but would follow as a natural extension.

Not only can this algorithm be used anywhere where a blood sample can be taken & there is computer access; it is able to provide more accurate results than a trained professional. It's marginal cost is zero, & does not warrant medical expertise past the ability to draw blood.

The dashboard available at link allows one to run predictions on the validation set, here are 5 sample predictions. One can see the prediction ('parasitized' or 'uninfected'), the model's confidence in the prediction & whether or not the prediction is accurate. Note: 4/5 of the predictions are accurate, whilst the inaccurate classification is ambiguous to the human I (parasites are not obviously visible) & the model took it's stance with far less conviction than in the other examples.


```
[[1]]
[1] "Parasitized"

[[2]]
[1] 0.99722

[[3]]
[1] "Accurately Classified"
```

```
[[1]]
[1] "Uninfected"

[[2]]
[1] 0.9836111

[[3]]
[1] "Accurately Classified"
```

```
[[1]]
[1] "Parasitized"

[[2]]
[1] 0.9986827

[[3]]
[1] "Accurately Classified"
```

```
[[1]]
[1] "Uninfected"

[[2]]
[1] 0.9947206

[[3]]
[1] "Accurately Classified"
```

```
[[1]]
[1] "Uninfected"

[[2]]
[1] 0.8329843

[[3]]
[1] "Misclassified"
```

References

1. khan academy - Accessed June 15, 2020. https://www.khanacademy.org
2. World Health Organization
3. Mandal, S., Sarkar, R.R. & Sinha, S. Mathematical models of malaria - a review. *Malar J* **10,** 202 (2011). https://doi.org/10.1186/1475-2875-10-202
4. WHO - malaria eradication
5. YANG, Hyun M. Malaria transmission model for different levels of acquired immunity and temperature-dependent parameters (vector). *Rev. Saúde Pública* [online]. 2000, vol.34, n.3 [cited 2020-06-18], pp.223-231. Available from: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102000000300003&lng=en&nrm=iso>. ISSN 1518-8787. http://dx.doi.org/10.1590/S0034-89102000000300003.
6. YANG, Hyun M and FERREIRA, Marcelo U. Assessing the effects of global warming and local social and economic conditions on the malaria transmission. *Rev. Saúde Pública* [online]. 2000, vol.34, n.3 [cited 2020-06-18], pp.214-222. Available from: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102000000300002&lng=en&nrm=iso>. ISSN 1518-8787. http://dx.doi.org/10.1590/S0034-89102000000300002.
7. Rajaraman S, Antani SK, Poostchi M, Silamut K, Hossain MA, Maude, RJ, Jaeger S, Thoma GR. (2018) Pre-trained convolutional neural networks as feature extractors toward improved Malaria parasite detection in thin blood smear images. PeerJ6:e4568 https://doi.org/10.7717/peerj.4568
8. Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Diagram 1: WHO Global Technical Strategy for Malaria

Diagram 2. Convolutional Neural Network Architecture

```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 30, 30, 32)        896

max_pooling2d (MaxPooling2D) (None, 15, 15, 32)        0

conv2d_1 (Conv2D)            (None, 13, 13, 64)        18496

max_pooling2d_1 (MaxPooling2 (None, 6, 6, 64)          0

conv2d_2 (Conv2D)            (None, 4, 4, 128)         73856

max_pooling2d_2 (MaxPooling2 (None, 2, 2, 128)         0

flatten (Flatten)            (None, 512)               0

dropout (Dropout)            (None, 512)               0

dense (Dense)                (None, 128)               65664

dense_1 (Dense)              (None, 1)                 129
=================================================================
Total params: 159,041
Trainable params: 159,041
Non-trainable params: 0
```

FIN.