


Fall 12-16-2016

Mechanistic Plug-And-Play Models for Understanding the Impact of Control and Climate on Seasonal Dengue Dynamics in Iquitos, Peru

Nathan Levick

Follow this and additional works at: https://digitalrepository.unm.edu/math_etds

 Part of the [Applied Mathematics Commons](#), [Mathematics Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Levick, Nathan. "Mechanistic Plug-And-Play Models for Understanding the Impact of Control and Climate on Seasonal Dengue Dynamics in Iquitos, Peru." (2016). https://digitalrepository.unm.edu/math_etds/90

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Mathematics & Statistics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Nathan Levick

Candidate

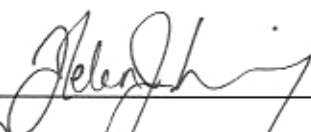
Mathematics & Statistics

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Helen Wearing



, Chairperson

Daniel Appelo



Yan Lu



**MECHANISTIC PLUG-AND-PLAY
MODELS FOR UNDERSTANDING THE
IMPACT OF CONTROL AND
CLIMATE ON SEASONAL DENGUE
DYNAMICS IN IQUITOS, PERU**

by

Nathan Levick

B.A., Fine Arts, University of New Mexico, 2010

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
Mathematics

The University of New Mexico
Albuquerque, New Mexico

December, 2016

Acknowledgments

First and foremost I would like to thank Dr. Helen Wearing, my advisor and thesis chair, for her support and guidance throughout the entire thesis timeline. She is truly a wonderful researcher, teacher, mentor, and person. I am grateful to my thesis committee, Dr. Yan Lu and Dr. Daniel Appelö, for taking the time to review and evaluate my work, as well as for the opportunity to learn from them in the classroom and take advantage of the devotion they bring to their students. Additionally, I would like to recognize all of my professors over the course of my mathematics post-baccalaureate and graduate career. I have made it to this moment through each of their contributions. I also acknowledge Dr. Michael Robert for his regular availability and assistance, especially in the early phases of modeling. And, largely through his organization, I want to thank the Wearing Lab and its members for enhancing my critical reading skills of scientific literature, providing essential feedback on my thesis manuscript and defense presentation as well as enriching my overall graduate school experience. Of course, I am very thankful to my friends and family for their patience and encouragement at every stage of my academic career. Except, I guess, for Luna who made it known that she would have much rather preferred that I was playing with her than working during those long hours at my desk.

MECHANISTIC PLUG-AND-PLAY MODELS FOR UNDERSTANDING THE IMPACT OF CONTROL AND CLIMATE ON SEASONAL DENGUE DYNAMICS IN IQUITOS, PERU

by

Nathan Levick

B.A., Fine Arts, University of New Mexico, 2010

M.S., Mathematics, University of New Mexico, 2016

Abstract

Dengue virus is a mosquito-borne multi-serotype disease whose dynamics are not precisely understood despite half of the world’s human population being at risk of infection. A recent dataset of dengue case reports from an isolated Amazonian city—Iquitos, Peru—provides a unique opportunity to assess dengue dynamics in a simplified setting. Ten years of clinical surveillance data reveal a specific pattern: two novel serotypes, in turn, invaded and exclusively dominated incidence over several seasonal cycles, despite limited intra-annual variation in climate conditions. Together with mechanistic mathematical models, these data can provide an improved understanding of the nonlinear interactions between the environmental and biological factors underlying dengue transmission as well as aid in the prediction of future epidemics. To examine the drivers of dengue in Iquitos we develop several stochastic discrete-time models and use likelihood-based plug-and-play inference techniques to explore potential factors that may explain the seasonal transmission pattern. By including climate-informed variables and accounting for known vector control measures in our model, we illustrate scenarios that can replicate the observed data and uncover the contribution of previously overlooked factors, such as the role of disease importation from human population migration. We discuss the implications of these results for understanding dengue dynamics in other endemic settings.

Contents

List of Figures	vii
List of Tables	viii
Glossary	ix
1 Introduction	1
1.1 Overview	1
1.2 Dengue	3
1.3 Infectious Disease Modeling	4
1.4 Inference for Nonlinear Dynamical Systems	5
1.5 Expected Results	6
2 Methods	8
2.1 Data	8
2.2 Model Structure	9
2.3 Theoretical Background	15
2.4 Analyses	20
3 Results	22
3.1 Preliminary Investigation	22
3.2 Temperature-Dependent EIP Model	23
3.3 Control Measures Model	24

Contents

3.4	Adjusted Vector Force of Infection Models	24
3.5	Parameter and Model Comparison	26
4	Discussion	28
4.1	Drivers of Seasonal Transmission	28
4.2	Modeling Process	30
4.3	Future Work	31
	Figures	33
4.4	Control _{$\lambda_{G,H}$} Plots	46
4.5	Control _{b_0} Plots	48
	Tables	50
	Bibliography	54
	Appendices	60

List of Figures

1	Case reports in Iquitos, Peru	33
2	Average temperature in Iquitos, Peru	34
3	Host-vector model	35
4	Schematic diagram of POMP model	36
5	Initial susceptible human population and average seasonal value . .	37
6	EIP likelihood profiles	38
7	EIP simulations	39
8	EIP model background infection	40
9	Control likelihood profiles	41
10	Control simulations	42
11	Control internal state simulations	43
12	Control _{λ_G} likelihood profiles	44
13	Control _{λ_G} Simulations	45
14	Control _{$\lambda_{G,H}$} likelihood profiles	46
15	Control _{$\lambda_{G,H}$} simulations	47
16	Control _{b_0} likelihood profiles	48
17	Control _{b_0} simulations	49

List of Tables

1	Summary of citywide fumigation efforts	50
2	Parameter list for demographic model	51
3	Altered Control Force of Infection hypotheses summary	52
4	Maximum point estimates for model parameters	52
5	Model comparisons: profile log-likelihood and AIC scores	53
6	Equation list for demographic model	62
7	Mathematical notation for methods in pomp constructor.	66

Glossary

\mathbb{N}_0	Natural numbers including zero, $\{0, 1, 2, \dots\}$.
$\mathbb{P}[X_n]$	Probability of random variable X at time t_n .
$X_{n,j}^{P,m}$	Random variable X at time t_n with particle j under distribution function P on iteration m .
$\mathcal{L}(\theta)$	Likelihood function of parameter vector θ .
$\ell(\theta)$	Log-likelihood function.
$\hat{\ell}(\theta)$	Estimate of log-likelihood function.
$\hat{\theta}$	Maximum Likelihood Estimate parameter vector.

1. Introduction

1.1 Overview

Recent increases in interconnectivity between human populations worldwide have intensified the public health threat of infectious diseases. Globalization trends like regular international travel and the effects of climate change have contributed to the geographic expansion and influence of several mosquito-borne viruses [1, 2], including dengue, chikungunya, West Nile, Japanese encephalitis, and Zika [3, 4]. In particular, dengue fever is the most rapidly spreading vector-borne disease (VBD), increasing 30-fold in the past five decades [5]. Dengue is currently endemic in over 100 countries [6], continuing to both emerge in those previously unaffected and reemerge in those formerly controlled [7]. The disease places an estimated 3 billion inhabitants of the world's tropical areas and approximately 120 million travelers at risk of acute infection each year while infecting 50–100 million with roughly 24,000 annual deaths [8, 9]. There is a significant economic burden from the attempts to prevent, treat, and control VBDs. For dengue alone, the associated cost of illness in the Americas is estimated at \$2.1 billion per year on average (2010 US dollars) [8].

Despite growing incidence and costs, current research lacks a precise understanding of the factors that contribute to dengue outbreaks. Mathematical modeling of infectious diseases can provide insight into the effectiveness of governmental policies, public health responses, and economic resources to mitigate and eliminate future occurrences. But, modeling efforts have not fully been able to predict what circum-

Chapter 1. Introduction

stances and mechanisms guarantee an outbreak. Dengue transmission is a complex phenomenon to model, the result of interactions between environmental variables and mosquito and human populations. Adding to the difficulty is a dearth of available and accurate, disaggregated data for parameterization and evaluation, especially for naive populations.

The Amazonian city Iquitos, Peru provides a prime setting for exploring the drivers of dengue outbreaks. Laboratory-confirmed dengue fever case reports from multiple health facilities in Iquitos collected over 10 years show that, despite an established history of dengue endemicity, a non-local strain of dengue virus invaded and dominated incidence over several seasonal cycles. Using these data, Stoddard et al. [10] compare dengue incidence to climatic variables to examine how the persistent, limited intra-annual variation of the climate conditions contribute to the seasonal pattern of dengue transmission through the year. Their descriptive and time-series analysis revealed a lack of conclusive evidence for a general interpretation of the temporal relationship between climate and dengue. This absence of a clear climatic driver suggests that a more nuanced model is necessary to complete the picture of Iquiteño dengue transmission dynamics.

Here, we construct a stochastic, compartmental host-vector model to analyze which drivers best contribute to the seasonal dengue pattern in Iquitos. Specifically, using a mechanistic framework we investigate the impact of control measures and climatic variables. We use a neoteric method of fitting mechanistic models to time-series data that optimizes input parameters using likelihood and only requires simulation from a dynamic model without the need for explicit transition probabilities, i.e. the “plug-and-play” property. With informed parameters, we are able to more directly hypothesize about the role of dengue incidence drivers in Iquitos. Our results suggest that control measures on the vector population and the importation of infected humans play an active role in dengue transmission in Iquitos, while some temperature-influenced parameters alone cannot adequately explain seasonal patterns.

1.2 Dengue

Dengue fever is a mosquito-borne viral disease common throughout tropical and subtropical regions [11], and, as aforementioned, its reach is expanding. Transmission of the virus cycles between humans and the predominate vector, female *Aedes aegypti* mosquitoes [12]. Symptomatic infection in humans shows as dengue fever. Lasting around a week, it produces flu-like symptoms, such as aching in the head, muscles and joints, and glands. There are four distinct but related dengue serotypes of the virus (DENV-1,2,3, and 4) [10]. Exposure to infection provides lifelong immunity to that particular serotype [13] with only partial and temporary cross-immunity to the others [14]. Hence repeated infection, by another serotype, is possible and can cause an intensified symptomatic response [7]. Severe dengue, or Dengue Hemorrhagic Fever (DHF), can be fatal if not properly recognized and treated. Although a dengue vaccine has very recently been licensed in a few countries for adults and children over 9 years old [15], there remains no long-term anti-viral drug solutions for dengue fever [16]. As a result, intravenous rehydration therapy is the primary clinical treatment [17], while vector control remains the best defense against transmission despite several challenges, including insecticide resistance, biosafety concerns, and regular community practice [5, 17].

Many models have been utilized to study the mechanisms that contribute to dengue transmission and outbreaks. Previous long-term studies of dengue in several endemic regions indicate intra-annual and inter-annual transmission patterns with locally varying drivers [18–20]. There are, however, also some commonalities independent of location. Short-term serotype cross-protection exists and impacts strategies for designing vaccine studies for multi-strain disease systems [14]. Climate is another known factor that affects dengue transmission, especially shaping the mosquito development and the virus replication processes [21]. Temperature influences larvae development rates, mortality, reproductive behavior, and the time it takes the virus

to disseminate in a mosquito, rendering the mosquito infectious, i.e. the extrinsic incubation period (EIP) [22–26]. Precipitation can also contribute to the mosquito lifecycle because standing water is the habitat for larvae and pupae [27].

1.3 Infectious Disease Modeling

Mathematical models of infectious disease help interpret observed epidemiological trends in data, which enables a better understanding of the critical underlying factors that lead to outbreaks and further transmission [28, 29]. While all models make assumptions about the relationship between variables, the statistical model aims to best describe data without trying to explain the reasoning behind variable interactions [30]. In the search for potential drivers of disease, statistical analysis can often find correlation between incidence and other seasonal patterns even without evidence of biological relevance [10]. A mechanistic approach, on the other hand, formulates the underlying equations based on observation and knowledge of the phenomenon’s behavior, necessitating the basis of model selection on scientific consideration over statistical convenience [31]. Compartmental dynamical systems are mechanistic models whose processes flow at prescribed rates between compartments over time. Common in epidemiology, compartmental models create a framework to describe how diseases move through populations, from susceptibility to recovery. In this way, the mechanistic model can quantitatively test and compare precisely expressed ideas about the driving factors behind a time series signal in the pursuit of finding the most explanatory correspondence between a model’s simulated output and the observed data.

Infectious disease is a complex, real-world phenomenon subject to nonlinearities and randomness. Biological population models, although simplifications, must be able to capture the process noise inherent in the data signal. Determinism assumes that processes follow an average rate ignoring variation within the mean. But, with the ability to capture inherent process randomness via rates defined by probability

distribution realizations, stochasticity attains and incorporates essential dynamic interactions commonly omitted by a purely deterministic skeleton [32]. Unpredictability in the timing of births, deaths, and interactions between individuals in a population is called demographic stochasticity. Environmental stochasticity is random variability in the conditions in which a system operates [31] and includes the external stochastic behavior that the model structure may not support or recognize. Failure to properly incorporate environmental stochasticity can lead to biases in parameter estimation effectually skewing the interpretation of results [33]. Thus, it is important to consider both forms of stochasticity when building the most rigorous model.

1.4 Inference for Nonlinear Dynamical Systems

Fitting mechanistic models to time series data is a challenging and active area of research [34–36]. Even simple models commonly found in the study of disease dynamics can not only be stochastic but also highly nonlinear, nonstationary, prone to measurement error and latent variables, or formulated in continuous time when data is sampled in discrete and perhaps irregular intervals [33]. The partially observed Markov process (POMP) model is one such model class that reconciles this assortment of requirements and is the focus of a new likelihood-based inference approach to data fitting through the optimization of model input parameter estimates.

A POMP, also commonly known as a hidden Markov or a state space model, involves an unobserved Markov state process (process model) and an explicit observational process (measurement model) [37]. State space models have applications across many domains that require making noisy and incomplete observations, e.g. economics, biology, and neuroscience [38]. A special brand of inference techniques, called plug-and-play, simulation-based, or equation-free, is a class of algorithms that can be applied to any time series data. They require only simulated samples from a model: closed-form expressions for transition probabilities are not necessary [31]. In other words, the mechanistic model becomes a “black box” where input parameters

turn into sample paths. Then, the exported paths simply plug into the inference machinery [39]. Since plug-and-play inference removes restrictions on model form, it opens up the scope of hypotheses that can be postulated as any applicable model can be inserted and tested, at the (potential) cost of computational effort [33].

For a mechanistic model of an epidemiological system, most parameters have an interpretable meaning. Maximum-likelihood via iterated filtering (MIF) estimation is a statistically efficient (no additional bias and variance in the resulting parameter estimates) [33] approach to fitting a model to data developed in response to these new types of dynamical system requirements [31]. MIF is a data-driven method of optimizing the model parameters, providing the best fit based on the provided observed data. By finding the most statistically consistent parameters for a soundly representative model, these meaningful parameter values will yield insights into the underlying dynamics of the phenomenon. Such likelihood-based methods are particularly effective in disease modeling. Previous topics of study include measles [33], pertussis [40], polio [41], pneumonia [42], and Ebola [43].

1.5 Expected Results

Iquitos, Peru presents an ideal opportunity to study the drivers of dengue transmission. First, Iquitos harbors the proper meteorological conditions for year-round mosquito presence, promoting arboviral disease endemicity. The annual persistent presence of disease provides a distinct domain to explore why and how epidemics emerge. Additionally, the epidemics we will study are due to a newly introduced serotype in the region. Gaining a better understanding of how a previously absent serotype of dengue virus is transmitted in a new territory can yield important considerations in the effort to curtail future spread. Notably, Iquitos has limited intra-annual variation in its meteorological conditions, yet it is not clear why there is a shifting, seasonal pattern of epidemics over many years. With reduced variability in meteorological conditions, historically common dengue drivers, alternative contributing

Chapter 1. Introduction

factors to disease incidence may be more discernible.

This thesis seeks to understand which factors contribute to the inter-annual variation in timing and magnitude of Iquitos' dengue epidemics. We develop a stochastic, discrete-time mechanistic model that reasonably reproduces the Iquitos dengue incidence data over the period of dominance for the single DENV-3 serotype, from its initial introduction to the region in 2001 to its replacement by the next dominating dengue serotype, DENV-4, in 2009. By using primed parameters derived from the iterative filtering approach of plug-and-play inference in a POMP-type dynamic system, we will investigate the impact of control and temperature-informed variables on the mosquito population dynamics. Were the control measures taken during this period effective or contributive to the patterns that resulted? Is the small variation in temperature enough to affect the overall transmission outcome? Are there any overlooked factors that gain importance as a result of this study? By addressing these questions, this model can add to the general understanding of dengue transmission in the ongoing mission of preventative public health.

2. Methods

2.1 Data

Iquitos, Peru is a mid-sized city with a 2012 population of 457,865. It sits isolated in the Great Plains of the Amazon drainage basin. Surrounded by the Amazon, Nanay, and Itaya rivers, Iquitos is not connected to a road system, making it only reachable by air and river travel. In 2000, the U.S. Naval Medical Research Center Detachment (NMRCDC) and the local Ministry of Health in Peru implemented a clinic-based syndrome surveillance system to investigate human febrile illness due to arboviruses by monitoring health clinics at several locations around Iquitos [44]. As a result, the study produced a ten-year dataset of weekly dengue incidence categorized by serotype (Figure 1).

While dengue has been continuously present in Iquitos since it was first reported in 1990, unlike many other dengue-endemic regions where multiple serotypes co-circulate, here it persists largely through repeated single serotype dominated invasions [45]. During the period of study, DENV-3 takes over from DENV-1 in early 2002, persisting almost exclusively until it is effectively replaced by DENV-4 around 2009. In addition, the data showcase a strong epidemic seasonality with peaks occurring during the height of summer, albeit with irregularity in the magnitude and timing. Inter-epidemic intervals last between 8 and 16 months based on wavelet analysis [10]. Figure 1 also highlights the five citywide control efforts that transpired during the period of study. Houses in specific city sectors were sprayed with three different ultra low vol-

ume (ULV), non-residual insecticides, which target adult mosquitoes. Each control period varied in length and the number of houses sprayed. Overall, approximately 80,000 different houses were treated (Table 1).

Iquitos experiences an equatorial climate with year-round rainfall and a limited range in temperature variation. Figure 2 shows the average daily temperature largely between 20 to 30 °C (68 to 86 °F) based on recordings from the Iquitos airport U.S. National Ocean and Atmospheric Administration (NOAA) weather station.

2.2 Model Structure

We build a discrete-time compartmental model that describes the transmission of dengue between human hosts and female *Ae. aegypti* mosquito vector populations utilizing an adaptation of a Reed-Frost chain binomial system to incorporate demographic stochasticity [29, 46]. By compartmentalizing based on the state of infection, we can investigate and capture the driving mechanisms in both populations (Figure 3). We define all probabilities on a daily time interval, from $[t, t + 1]$. Individuals enter and leave population compartments via draws from governing transitional probabilities.

The human population (Figure 3a) divides into susceptible, exposed (infected but not infectious), infectious, and removed (quarantined, recovered, or immune) classes, represented as $S_H(t)$, $E_H(t)$, $I_H(t)$, $R_H(t)$, respectively. Let the total human population at time t be $N_H(t) = S_H(t) + E_H(t) + I_H(t) + R_H(t)$. Similarly the compartmentalization of the vector population (Figure 3b) divides into juvenile (preliminary developmental stage) alongside adult susceptible, exposed, and infectious classes, $J_G(t)$, $S_G(t)$, $E_G(t)$, $I_G(t)$, respectively. The total vector population at time t is $N_G(t) = S_G(t) + E_G(t) + I_G(t)$. Note, the brevity of the average lifespan of a vector precludes the inclusion of a recovery class. See Table 2 for a list of the parameters used in the equations below.

Human Dynamics

We use one of two versions of the daily probability of a susceptible human becoming exposed

$$\lambda_H(t) = 1 - \exp\left(-\frac{\beta(t)I_G(t)}{N_H(t)}\right) \quad (1)$$

$$\lambda_H(t) = 1 - \exp\left(-\frac{b_0 I_G(t)}{N_H(t)}\right) \quad (2)$$

depending on how much emphasis seasonality should play into the exposure of the human population. Here $\beta(t)$ is the transmission rate. A simplified cosine captures a general form of seasonal transmission

$$\beta(t) = b_0 \left[b_1 \cos\left(2\pi\left(\frac{t}{365} + b_2\right)\right) + 1 \right]. \quad (3)$$

The parameters b_0 , b_1 , and b_2 determine the shape of the curve: average value, amplitude, and shift, respectively, where $b_0 \in \mathbb{R}^+$ and $b_1, b_2 \in [0, 1]$. Note the only difference from Equation (1) to Equation (2) is $b_1 = b_2 = 0$. The probability of a host becoming infectious, σ_H , is defined in terms of the average intrinsic incubation period, $(\hat{\sigma}_H)^{-1}$

$$\sigma_H = 1 - \exp(-\hat{\sigma}_H). \quad (4)$$

The probability that an infectious host recovers, γ_H , is defined in terms of the human average duration of infectiousness, $(\hat{\gamma}_H)^{-1}$:

$$\gamma_H = 1 - \exp(-\hat{\gamma}_H). \quad (5)$$

As the period of interest covers almost a decade, births and deaths of humans and mosquitoes are included in the demography setup. We converted annual births and deaths per 1000 people for Peru in 2014 to a daily rate of a person of being born and dying, given by ϕ_H and μ_H respectively. In total, the human population is described

Chapter 2. Methods

by the following transition probability definitions and equations:

$$\begin{aligned}
 B_H(t) &\sim \text{Poisson}(N_H(t)\phi_H) \\
 [W_H(t), X_H^S(t)] &\sim \text{Multinomial}(S_H(t), (1 - \mu_H)\lambda_H, \mu_H) \\
 [V_H(t), X_H^E(t)] &\sim \text{Multinomial}(E_H(t), (1 - \mu_H)\sigma_H, \mu_H) \\
 [U_H(t), X_H^I(t)] &\sim \text{Multinomial}(I_H(t), (1 - \mu_H)\gamma_H, \mu_H) \\
 X_H^R(t) &\sim \text{Binomial}(R_H(t), \mu_H)
 \end{aligned}$$

$$\begin{aligned}
 S_H(t+1) &= S_H(t) + B_H(t) - W_H(t) - X_H^S(t) \\
 E_H(t+1) &= E_H(t) + W_H(t) - V_H(t) - X_H^E(t) \\
 I_H(t+1) &= I_H(t) + V_H(t) - U_H(t) - X_H^I(t) \\
 R_H(t+1) &= R_H(t) + U_H(t) - X_H^R(t).
 \end{aligned}$$

Above $B_H(t)$, $W_H(t)$, $V_H(t)$, and $U_H(t)$ represent the number of newly susceptible, newly exposed, newly infectious, and newly recovered individuals, respectively, at time t . Likewise, the respective $X_H^*(t)$ signifies the number of deaths from each category, (S, E, I, R) .

Vector Dynamics

The vector population dynamics follow a similar setup but with differences reflective of how dengue interacts with the *Ae. aegypti* lifecycle. We use a simple mosquito age structure to inform entry into the S_G class. The mosquito lifecycle begins with eggs that hatch in water and undergo a “juvenile” stage, growing from larvae to pupae, before emerging as adult insects. Recruitment begins with the parameters $\hat{\phi}_G$, the daily average adult female mosquito egg production rate, and $\hat{\theta}_G$, the average rate of egg survivorship. The probability of maturation, ρ_G , based on the average juvenile vector development duration, $(\tau)^{-1}$, is defined as

$$\rho_G = 1 - \exp(-\tau). \quad (6)$$

Chapter 2. Methods

The force of transmission is slightly changed in the vector probability equation to include the demographic importation of humans from outside of Iquitos into the city and environmental stochasticity through a multiplicative noise term. Again, there are two variations of the probability of infection. The first is

$$\lambda_G(t) = 1 - \exp\left(-\frac{\beta(t)}{N_H(t)}(I_H(t) + \eta)\zeta(t)\right) \quad (7)$$

where η is the average daily human importation of infections and $\zeta(t)$ is Gamma white noise with intensity ϵ . Homogeneous mixing is assumed. To capture a seasonal η effect, the second version defines the probability of vector infection as

$$\lambda_G(t) = 1 - \exp\left(-\frac{(b_0 I_H(t) + \beta(t)\eta)\zeta(t)}{N_H(t)}\right). \quad (8)$$

After exposure, the probability that a vector becomes infectious, σ_G , is defined in terms of the average extrinsic incubation period (EIP), $(\hat{\sigma}_G)^{-1}$:

$$\sigma_G = 1 - \exp(-\hat{\sigma}_G). \quad (9)$$

There is empirical evidence to suggest that average EIP follows an exponential decay with temperature based on the mechanics of virus replication in the mosquito. Hence, in some model scenarios, we use a temperature-dependent EIP, such that

$$\sigma_G(T) = 1 - \exp(-\hat{\sigma}_G(T)) = 1 - \exp\left(-\frac{1}{c_1 \exp(-c_2 T)}\right) \quad (10)$$

where covariate $T = T(t)$ is the daily average temperature from Figure 2 and c_1 and c_2 are positive real function shape parameters. Death rates differ between the juvenile and adult vector classes. Daily juvenile deaths are given by

$$\mu_J(t) = 1 - \exp(-\hat{\mu}_J - \alpha J_G(t)) \quad (11)$$

with a background mortality rate, $\hat{\mu}_J$, calculated to provide 80% survivorship (§A.2), and strength of juvenile competition factor, α . Equilibrium analysis performed for α assumes an initial mosquito population taken at a ratio of approximately 2:1 compared

Chapter 2. Methods

to humans in Iquitos (§A.3). Meanwhile, the probability of daily adult mortality, μ_G , is defined based on the average adult vector lifespan, $(\check{\mu}_G)^{-1}$

$$\mu_G = 1 - \exp(-\check{\mu}_G). \quad (12)$$

A separate control measure scheme alters μ_G to investigate the effects of control measures during this time frame. The periods of control are given a higher probability of daily adult mortality, weighted by the efficacy of the fumigation, i.e. the number of sprayed houses per duration of each control period. As such, the probability of daily adult mortality during control is defined based on a weight w_i for control period i and a new weighted average lifespan parameter, $(\hat{\mu}_G)^{-1}$

$$\mu_G = 1 - \exp(-w_i \hat{\mu}_G). \quad (13)$$

Finally, using these probabilities, the vector population dynamics are described by the following:

$$\begin{aligned} B_G(t) &\sim \text{Pois}(\hat{\phi}_G \hat{\theta}_G N_G(t)) \\ [Z_G(t), X_G^J(t)] &\sim \text{Multinom}(J_G(t), (1 - \mu_J(t))\rho_G, \mu_J(t)) \\ [W_G(t), X_G^S(t)] &\sim \text{Multinom}(S_G(t), (1 - \mu_G)\lambda_G(t), \mu_G) \\ [U_G(t), X_G^E(t)] &\sim \text{Multinom}(E_G(t), (1 - \mu_G)\sigma_G, \mu_G) \\ X_G^I(t) &\sim \text{Binom}(I_G(t), \mu_G) \end{aligned}$$

$$\begin{aligned} J_G(t+1) &= B_G(t) - Z_G(t) - X_G^J(t) \\ S_G(t+1) &= S_G(t) + Z_G(t) - W_G(t) - X_G^S(t) \\ E_G(t+1) &= E_G(t) + W_G(t) - U_G(t) - X_G^E(t) \\ I_G(t+1) &= I_G(t) + U_G(t) - X_G^I(t) \end{aligned}$$

Similarly, as before, $B_G(t)$, $Z_G(t)$, $W_G(t)$, and $U_G(t)$ represent the number of newly juvenile, newly susceptible, newly exposed, and newly infectious individuals, respec-

tively, at time t . Additionally, respective $X_G^*(t)$ signifies the number of deaths from each category, (J, S, E, I) .

Measurement Model

The observed data represents only a sample of the total dengue incidence in the population during the targeted period: those identified and recorded by the surveillance system. The measurement feature of the state space model mirrors the reporting process by subselecting a number of cases from all infections. The model uses an overdispersed binomial distribution to build in higher than expected variance into the reporting process to account for underreporting and general measurement error that naturally occur in the precise surveillance that produced the data. We let c_H be the number of weekly case reports selected from the weekly total of true incidence C , a sum of seven daily changes in U_H , the surviving infected population en route to recovery. The amount of case reports follows from

$$c_H(t) | C \sim \text{Normal}(\rho C, \rho(1 - \rho)C + (\psi \rho C)^2) \quad (14)$$

where $\rho \in [0, 1]$ is the reporting rate efficiency, and ψ is the overdispersion factor. Note that $\psi = 0$ produces the binomial case. The sampling of the measurement modeling is described in §A.4.

Reproductive Numbers

In epidemiology, there are standard metrics that describe the potential for disease to spread. The basic reproductive number, R_0 , is the average number of secondary cases a typical infected case generates in a completely susceptible population. For this discrete model (neglecting human demography in the exposed and infectious periods), the reproductive number is approximately defined [29] to be

$$R_0(t) = \left(\frac{N_G(t)\beta_H\beta_G}{N_H(t)\gamma_H\mu_G} \right) \left(\frac{\sigma_G(1 - \mu_G)}{\sigma_G(1 - \mu_G) + \mu_G} \right) \quad (15)$$

where

$$\beta_H = \begin{cases} \beta(t), & \text{for (1)} \\ b_0, & \text{for (2)} \end{cases} \quad \text{and} \quad \beta_G = \begin{cases} \beta(t), & \text{for (7)} \\ b_0, & \text{for (8)} \end{cases} \quad (16)$$

depending on the probability of exposure assumptions. The effective reproductive number, R_f , takes into account changes within the susceptible population and is defined as

$$R_f(t) = \left(\frac{S_H(t)}{N_H(t)} \right) R_0(t). \quad (17)$$

Values of $R_f > 1$ indicate a growing infectious population and, conversely, $R_f < 1$ signifies a declining infectious population.

2.3 Theoretical Background

Partially Observed Markov Process (POMP) Models

A POMP models data as noisy and incomplete observations of random variables defined by a stochastic process known as a Markov chain $\{X_n, n \in \mathbb{N}_0\}$. A process is considered Markovian if given the past states X_0, X_1, \dots, X_{n-1} and present state X_n , the conditional distribution of any future state X_{n+1} is only dependent on the present state X_n and independent of the past states. From the set $T \subset \mathbb{R}$, let a time index be defined as $\{t_i \in T, i = 1, \dots, N\}$ with initial time $t_0 \in T$ and $t_0 \leq t_1 < \dots < t_N$. Then let $X_{0:N} = (X_0, \dots, X_N)$, where $X_n = X(t_n)$, a latent dynamic process, be such an unobserved Markov chain. In our case, in which the state variables represent a vector of various population sizes from the compartmental model, values are in the state space $\mathbb{X} \subset \mathbb{N}_0^{\dim(\mathbb{X})}$. The process $X_{0:N}$ is unobserved and only realized by another process $Y_{1:N} = (Y_1, \dots, Y_N)$ with observable random variable Y_n valued in the observation space $\mathbb{Y} \subset \mathbb{N}_0^{\dim(\mathbb{Y})}$ at time t_n . The data $y_{1:N}^* = y_1^*, \dots, y_N^*$ is a sequence of N fixed observations of $X_{1:N}$ from $Y_{1:N}$ at the corresponding times $t_{1:N}$. In the model above, each Y_n is a single value, unlike each X_n , and derived from the overdispersed

binomial. Note that X_0 initializes the process model, but the first observation Y_1 occurs from X_1 . As part of a Markov chain, each state X_n is independent of all other states (in both the state and observation process) beside the one immediately preceding it,

$$\mathbb{P}[X_n|X_0, \dots, X_{n-1}, Y_1, \dots, Y_{n-1}] = \mathbb{P}[X_n|X_{n-1}]. \quad (18)$$

Meanwhile, each measurement Y_n for all $n = 1, \dots, N$ depends only on the state at the coinciding time,

$$\mathbb{P}[Y_n|X_0, \dots, X_n, Y_1, \dots, Y_{n-1}] = \mathbb{P}[Y_n|X_n]. \quad (19)$$

Figure 4 illustrates the relationship between the state and observational process and the inherent dependencies of the variables.

Likelihood for POMP Models

The Markov process, and inherently its states, also depend on the parameters of the mechanistic model. Let a “particle” refer to any specific combination of parameters, i.e. any vector θ in the parameter space $\Theta \subset \mathbb{R}^{\dim(\Theta)}$. (Note that in this thesis, vector notation for the parameter vector θ and for the process model states $X_{0:N}$ will not be used for clarity of exposition.) Through statistical inference it is possible to determine which values of θ , if any, can reasonably and most accurately represent the time series data. The likelihood function is a metric that measures goodness of fit to the data, defined as the probability of a given set of data D having occurred under a particular hypothesis H , i.e. $\mathcal{L}(H; D) = \mathbb{P}[D|H]$. In this case, the data is represented by the time series realization $y_{1:N}^*$, and the hypothesis is a particular set of parameters, θ .

Separate probability density functions describe the state and measurement random variables. Using continuous density distributions, if we combine the one-step conditional transition density from latent state to state, written as $f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta)$, and initial density, $f_{X_0}(x_0; \theta)$, with the conditional measurement density, $f_{Y_n|X_n}(y_n|x_n; \theta)$,

Chapter 2. Methods

then their joint distribution is

$$f_{X_{0:N}, Y_{1:N}}(x_{0:N}, y_{1:N}; \theta) = f_{X_0}(x_0; \theta) \prod_{n=1}^N f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta) f_{Y_n|X_n}(y_n|x_n; \theta). \quad (20)$$

Using Equation (20) and marginalizing, we form the likelihood function for a POMP:

$$\begin{aligned} \mathcal{L}(\theta) &= f_{Y_{1:N}}(y_{1:N}^*; \theta) \\ &= \int \prod_{n=1}^N f_{Y_n|X_n}(y_n^*|x_n; \theta) f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1}^*; \theta) dx_{0:N} \end{aligned} \quad (21)$$

$$= \prod_{n=1}^N f_{Y_n|Y_{n-1}}(y_n^*|y_{1:n-1}^*, \theta). \quad (22)$$

The explicit formula in Equation (21) is a high dimensional integral and, as is often the case, may not be known or difficult to express. Yet, it exists, is theoretically explicable as a likelihood function, and is numerically calculable through the inference algorithm. Furthermore, the log transformation of the likelihood, the log-likelihood, is more preferable to facilitate the ease of the maximization calculation in the algorithm described in the next section. From Equation (22), we find the log-likelihood to be represented as follows:

$$\begin{aligned} \ell(\theta) = \log \mathcal{L}(\theta) &= \log \prod_{n=1}^N f_{Y_n|Y_{n-1}}(y_n^*|y_{1:n-1}^*; \theta) \\ &= \sum_{n=1}^N \log f_{Y_n|Y_{n-1}}(y_n^*|y_{1:n-1}^*; \theta) \\ &= \sum_{n=1}^N \ell_{n|1:n-1}(\theta). \end{aligned} \quad (23)$$

Maximum Likelihood via Iterated Filtering (MIF)

Using the iterative filtering procedure, we can search for the best parameter set to input into our model that will most likely result in output shaped like the observed

Chapter 2. Methods

data signal. MIF specifically uses iterated filtering to search for the maximum likelihood estimate (MLE), the particle value $\hat{\theta}$ that maximizes $\ell(\theta)$ for all particles $\theta \in \Theta$. IF2 is a newly developed MIF algorithm (§B.1) that uses iterated, perturbed Bayes maps to improve computational performance over former iterated filtering implementations [47]. In pursuit of $\hat{\theta}$, the intuition behind IF2 is to take an initial “swarm” of J particles, $\{\Theta_j^0, j = 1, \dots, J\}$, and update each particle in the swarm via a random walk over each time step in the time series. The process reiterates the random walk M times, each time with a smaller variance, $\sigma_m, m = 1, \dots, M$, based on a decreasing function called the “cooling schedule.” The previous output swarm $\Theta_{1:J}^{m-1}$ is then input as the starting particles for the next iteration. This procedure theoretically converges to the region of Θ with maximal maximum likelihood [37].

Within the MIF process, the filtering occurs through the sequential Monte Carlo (SMC) algorithm, or the “particle filter,” which provides the standard numerical method to obtain the likelihood estimates for POMP [33]. SMC estimates the likelihood function from Equation (23) via iterated Bayes Maps and an approximation technique called Importance Sampling by reducing the high-dimensional integral to a sequence of lower dimensional problems [37]. The basic idea of the algorithm is a recursive random walk process in time, i.e. sequential, that iterates two steps, a prediction step and a filtering update step. In general at time step t_n for a single swarm particle j , the latent process state $X_{n,j}$ and particle $\Theta_{n,j}$ are first predicted from the previously found latent process state $X_{n-1,j}$ and particle $\Theta_{n-1,j}$ a step before at t_{n-1} . Then, weights for each particle j are calculated based on their proportional “importance,” i.e. the probability that the observed data point at t_n, y_n^* , resulted from the predicted sample point value X_n . Finally, the swarm particles and the latent process states are updated through a filtering, i.e. resampling, as fresh draws based on the new weighted probabilities of being close to y_n^* . The iterations continue until $n = N$ when the time series ends.

Looking in more detail at how MIF processes the state random variables, let

Chapter 2. Methods

$\{x_{n,j}^P\}_{j=1}^J$ be a sample of point estimates for the latent state drawn from the “prediction distribution” $X_n|y_{1:n-1}^*$ at t_n and $\{x_{n-1,j}^F\}_{j=1}^J$ be a set of point estimates drawn from the “filtering distribution” $X_n|y_{1:n}^*$ at time t_{n-1} . The prediction distribution, in general, is a plug-and-play simulation of the state to state process and defined as

$$X_{n,j}^P \sim f_{X_n|X_{n-1}}(x_n|x_{n-1,j}^F; \theta). \quad (24)$$

Note $f_{X_0}(x_0; \theta)$ informs the first prediction distribution as $x_{0,j}^F$. The weights are derived from an Importance Sampling procedure based on an importance density setup by a recursive relation between the time steps n and $n - 1$ that underlies the entire SMC algorithm. The weight equation is defined as the probability of simulating the observed data from the given predicted state value, written as

$$w_{n,j} = f_{Y_n|X_n}(y_n^*|x_{n,j}^P; \theta). \quad (25)$$

After the weights are defined from the predicted sample latent states $\{x_{n,j}^P\}$, these sample points are then filtered. J draws are taken from the probabilities based on a normalization of the weights found in Equation (25) and given reordered indices $k_{1:J}$ to become $\{x_{n,k_j}^P\}$, the new filtered set of latent state values to be used in the next time step iteration to inform the prediction step, i.e. $X_{n,j}^F = X_{n,k_j}^P$. As a result, the estimate of the log-likelihood Equation (23) can be found as follows:

$$\begin{aligned} \hat{\ell}(\theta) &= \sum_{n=1}^N \hat{\ell}_{n|1:n-1}(\theta) \\ &= \sum_{n=1}^N \log \left(\frac{1}{J} \sum_{j=1}^J \mathbb{P} \left[y_n^* | x_{n,k_j}^P; \theta \right] \right) \end{aligned} \quad (26)$$

$$= \sum_{n=1}^N \log \left(\frac{1}{J} \sum_{j=1}^J w_{n,k_j} \right). \quad (27)$$

In this way, the values closest to the data at each time step, and the model parameters that produce them, are rewarded with higher probability, and likelihood, of being realized while the values that are more distant are filtered out. To complete the

MIF process, the SMC algorithm is repeated M times, starting each iteration at the beginning of the times series. The smaller random walk variances, following the cooling schedule, further hone the filtering process.

2.4 Analyses

The data were restricted to the period between 1 October 2001 and 1 July 2009 to focus on the rise and fall of DENV-3. During a two week period in December 2004, at the height of an epidemic, the surveillance system in one hospital was extended to take advantage of the increased patient influx, resulting in increased reporting. To reconcile the difference in reporting from the other weeks, we rescaled the cases during these two weeks using the ratio of the maximum number of negative cases during the heightened reporting to that in the rest of the data.

Regarding numerical implementation, we utilized the R package `pomp.R` [39] to facilitate the simulation and analysis of a state space model constructed to fit the Iquitos incidence data. In the library, a POMP model is of class `pomp`, created with a detailed constructor (§B.2) and featuring four main parts. The simulation of state to state transition $f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta)$ occurs in the `rprocess`, and the `dprocess` provides an evaluation of this density. Likewise, the `rmeasure` section holds the procedure for a simulation of the state to measurement transition $f_{Y_n|X_n}(y_n|x_n; \theta)$ and the `dmeasure` takes a draw. Note that simulation-based methods require an `rprocess` but not a `dprocess` since there is no explicit function to evaluate. Other inputs to the constructor are the specified model classes, parameter names, initial values, covariates, and parameter constraints (to reduce the optimization search space). Also noteworthy, the `pomp` class allows for simulation on a finer time scale than the data. Our model represents a process that is best described on a daily time scale, whereas the data are case reports aggregated at a weekly time scale.

While `pomp.R` contains a suite of methods to solve a variety of state space problems, we focus on the implementation for a “full-information”, i.e. non-summary statistic-

Chapter 2. Methods

based likelihood function, plug-and-play, frequentist POMP model. The solution of this specific class of POMP model results from iterated filtering, represented by the IF2 algorithm command `mif2`. Inputs to `mif2` include the `pomp` object, the number of particles in the swarm, number of iterations, standard deviation of the random walk, and shape of cooling schedule.

To find the maximal parameters for each model, we first create a profile design containing the initial values for a range of particles to be optimized by MIF. Within the profile design a single parameter is selected to be “profiled,” or explored, over a certain range while the other parameters move freely in the random walk. Each initial particle plugs into a `mif2` command, outputting a point estimate particle with the maximal log-likelihood. For each of our `mif2` runs we use 50 iterations, 1000 filtering particles, random walk standard deviation of 0.02 for each parameter, and a geometric cooling schedule shape. At this stage, the single point estimate of likelihood produced is usually an approximation of $\hat{\theta}$, not sufficiently reliable for inference. Hence, each particle at the end of the `mif2` command goes through 10 additional particle filters, this time with 2000 particle-sized swarms, in order to polish the result, calculate a standard error, and ultimately give the most accurate picture of the likelihood surface. Finally, we pluck the top particles with the highest log-likelihood to create a new profile design and repeat the previous `mif2` and multiple particle filter progressions.

3. Results

3.1 Preliminary Investigation

Early simple model formulations revealed information regarding the nature of some common parameters. The shift in the transmission function, b_2 , failed to show significant difference from a zero value, indicating that the peak of the cosine transmission function closely aligned with 1 December, the approximate middle of summer. As a consequence, it was held constant at 0 and removed from subsequent particles. The initial distribution of the human population into the model compartments affects the model output. The initial susceptible population became a reoccurring particle parameter, and some experimentation indicated the initial human exposed and infectious populations should be greater than 1. Accordingly, we used $E_H(t_0) = E_0 = 6$ and $I_H(t_0) = I_0 = 2$ for all models. There was also a marked relationship between the average seasonal transmission value, b_0 , and the initial human population, $S_H(t_0) = S_0$ (Figure 5). The inverse relationship makes sense in the context of the effective reproductive number. Recall $R_f(t) = \left(\frac{S_H(t)}{N_H(t)}\right) R_0(t)$ and measures the potential for change in the infectious population over time. Because S_0 informs S_H and b_0 informs $\beta(t)$ and thus R_0 , to maintain the same R_f : if S_0 is low, b_0 compensates with a large value, and conversely, if S_0 is high, b_0 becomes small. Using these observations, we produced five models to test hypotheses about the driving factors of seasonal dengue transmission and ascertain the best fitting simulated signals representative of the original incidence data.

3.2 Temperature-Dependent EIP Model

The magnitude of seasonal variation in the area’s climate was modest but extant. A model with a daily average temperature-dependent EIP formulation from Equation (10), written as **EIP**, investigated if this small differentiation throughout the year contributed to the overall seasonality in transmission. A previous version of this model did not incorporate background infection. Consequently, the model simulations were unable to sustain infections beyond initialization. Therefore, to account for the re-introduction of infection, we added human importation via the parameter η . Then, with the particle $\theta = [S_0, \rho, \psi, b_0, b_1, c_1, c_2, \eta, \epsilon]$, we specifically profiled over 15 equispaced values of S_0 between 0 and 350,000. Likelihood plots of each parameter (Figure 6) show which values have the highest log-likelihood score based on all the particles tested. Note a local regression smoothed (LOESS) curve with 95% confidence region accompanies each parameter scatterplot. The initial susceptible human population reveals little information with relatively equal maximal log-likelihood values across the entire range. Profiles of c_1 and c_2 also highlight the wide ranges. Both EIP shape function parameters have particles that cluster together with some larger values complicating the identifiability of the parameter. The other parameters ρ , ψ , b_1 , and ϵ , however, demonstrate more identifiable definition, varying in a smaller value range.

The EIP function generated by the maximal likelihood shape parameters assigns the average EIP to be 0 days for the entire duration, meaning there is no latent period between exposure and infection for mosquitoes. The exposed compartment is effectively removed from the model; as soon as a mosquito is infected it is infectious. The model, despite the unrealistic EIP value, produces a high log-likelihood particle rating, -1078.19 . The associated simulations of the infectious human population over time from this optimal particle (Figure 8) exemplify how the infectious population is continually replenished, rarely dipping below the daily importation of

infectious humans, in this case $\eta = 40.61$. Clearly, transmission is bolstered by a large daily amount of external infectious humans and highly efficient transmission in the mosquitoes, a result of a zero EIP. Simulations of the R_0 curve (Figure 7b) give values well above the expected R_0 for dengue, 3 – 5.

3.3 Control Measures Model

The model denoted **Control** tests the impact of control measures taken during this period on transmission with particle $\theta = [S_0, \rho, \psi, b_0, b_1, \eta, (\hat{\mu}_G)^{-1}, \epsilon]$. In a profile over S_0 (Figure 9), the parameters ψ , b_1 , $(\hat{\mu}_G)^{-1}$, and ϵ have distinct ranges in which the best values arise. Specifically, the profile for $(\hat{\mu}_G)^{-1}$ indicates that during control periods the weighted average adult vector lifespan should be around 8 or 9 days, less than the non-control period, 10 days. Again, the S_0 values are fairly equal across the entire range of acceptable profiled values. The remaining parameters, ρ , b_0 , and η show strong bunching in a particular range but also possess relatively equal likelihood values beyond that cluster, in effect limiting their identifiability. The basic reproductive number (Figure 10b) illustrates the timing and overall effect of control. The periods under control lead to drops in the R_0 as a consequence of reductions in the mosquito population. The overall maximum log-likelihood for the control model is -1081.80 . Note although the likelihood scores are about equal, **Control** has one less degree of freedom than **EIP** as well as stronger interpretations of parameter results. Lastly, the optimal particle simulations of the internal states (Figure 11) portray the information tracked during each run and pictorially reinforce the notion that stochasticity promotes small but visible differences in output for each model run.

3.4 Adjusted Vector Force of Infection Models

The large estimated values of η resulting from both **EIP** and **Control** indicate that human importation is capturing a large portion of the transmission dynamics. To attempt to bring supplementary insight into the timing and seasonal variation of im-

ported infections, we proposed alternative implementations of a seasonal η through three model hypotheses, summarized in Table 3. Each of these models alters the **Control** by manipulating the Force of Infection (FOI) term in the probability of exposure, i.e. F such that $\lambda_*(t) = 1 - \exp(-F_*)$, in the human and mosquito populations. Altering the FOI changes the assumptions about how seasonality in the model, represented by the cosine transmission function, affects internal and external infections.

First, we hypothesized and implemented a simple seasonal human importation scheme in **Control** $_{\lambda_G}$. Starting from **Control**, this model substitutes an altered vector force of infection into the probability of exposure λ_G from Equation (8) so that the cosine transmission function affects external infectious humans but not internal infectious humans. The separation provides a mechanism for seasonal human importation. Another S_0 profile here reveals likelihoods with similar outputs as **Control** (Figure 12). The ranges for the parameters ρ , ψ , b_1 , $(\hat{\mu}_G)^{-1}$, ϵ and even η are mostly unchanged. The biggest difference among particle value trends between **Control** and **Control** $_{\lambda_G}$ occurs in a slight range extension for b_0 . Despite the similarity of profile output, the top particle reveals some differentiation, including a lowered ρ , 0.063 from 0.27, and larger η , 24.33 compared to 10.03. **Control** $_{\lambda_G}$ generates an improved profile log-likelihood score of -1080.84 from its top particle, creating the simulations in Figure 13a. Perhaps the most indicative result, an increase in the magnitude of R_0 (Figure 13b) beyond the usual 3 – 5 range suggests potential complications in the interpretability of the recovered particle values.

Using FOI modifications, we remove the current interference from the overlap between the host and vector populations, which might be causing the high R_0 readings from **Control** $_{\lambda_G}$. Correspondingly, in addition to changing λ_G to Equation (8), as done in **Control** $_{\lambda_G}$, the first variation **Control** $_{\lambda_G,H}$ also updated λ_H from Equation (1) to Equation (2), thereby removing seasonality from the force of human infection such that seasonal forcing only acts on external importation. The profiles for **Control** $_{\lambda_G,H}$ (Figure 14) show similar results to **Control** $_{\lambda_G}$ with no general improvement to param-

eter identifiability. In fact with the changes, b_1 loses definition altogether. The final modification **Control** $_{b_0}$ decouples the seasonality between the human and mosquito populations completely by using distinct average transmission parameters in λ_H and λ_G , denoted by b_0^H and b_0^G respectively. Specifically, the profiles for **Control** $_{b_0}$ (Figure 16) show that the values for b_0^H and b_0^G are distinct. In both of these models, the isolation of seasonality to external infections is apparent in the reproductive number curves (Figure 15b and Figure 17b), no longer exhibiting the cosine shape. Further, the values of R_0 reduce to a more appropriate range compared to **Control** $_{\lambda_G}$.

3.5 Parameter and Model Comparison

We can check the parameters for consistency by comparing across models. Table 4 shows the top scoring log-likelihood particle from each model. We see that parameters such as ψ and ϵ vary little across models. A parameter with consistency across models suggests evidence for maintaining a fixed value going forward. Other parameters like b_1 varies widely indicating a stronger impact in the differences between models. More moderation is found in the ranges of b_0 as well as η , which demonstrates a value above initial expectations.

In comparing between models, the log-likelihood score only accounts for goodness of fit and not the differences in model complexity. The Akaike information criterion (AIC) is a numerical means for model comparison that considers both, including complexity via the number of free parameters. For model i , AIC is defined based on the estimated log-likelihood calculated of the MLE from the IF2 algorithm and the particle length, p_i :

$$\text{AIC} = 2p_i - 2 \left(\hat{\ell}(\hat{\theta}) \right)_i. \quad (28)$$

The interpretation of AIC gives preference to the model with the smallest value. Positive evidence of a model being superior to another is given only when there is a difference in AIC of at least 2 between models.

Table 5 summarizes the log-likelihood and AIC scores of the optimal particle

Chapter 3. Results

for each model tested. **EIP** has the lowest AIC at 2174.38, yet the temperature-dependent EIP curve that resulted was zeroed out, producing unrealistic parameter values for the model since there is a known and existent EIP in the mosquito dynamics. The difference between **Control** at 2179.60 and **Control** $_{\lambda_G}$ at 2177.68 is just less than 2, indicating there is not enough evidence to declare preference of one over the other. However, the high value of the R_0 of **Control** $_{\lambda_G}$ indicates interpretability issues, signaling **Control** as an overall better model choice in this pair. Similarly, the difference between **Control** $_{\lambda_{G,H}}$ at 2182.20 and **Control** $_{b_0}$ at 2181.50 is also not substantial enough to differentiate the two models. While **Control** $_{\lambda_{G,H}}$ is the weakest model tested, **Control** $_{b_0}$ is not significantly different from **Control** based on AIC alone.

4. Discussion

4.1 Drivers of Seasonal Transmission

Dengue virus transmission dynamics in Iquitos exhibit a periodicity that suggests seasonal forcing amid a year-round presence. With mechanistic models we were able to directly investigate hypotheses about this pattern. The impact of control exemplifies the power of the plug-and-play apparatus. By including a control mechanism, through the addition of a single model parameter, we receive a strong indication that the resulting average adult vector lifespan is shorter during periods of control than during periods without control administered. The impact of control measures is often difficult to detect because control is usually applied mid-outbreak. Afterwhich, if incidence declines, it can be difficult to determine if the reduction was due to a depletion of susceptibles or due to the positive results of the control measure. Therefore, it is often argued to have little impact on transmission. Yet here, a simple control mechanism based on the fumigation record shows clear influence.

These positive results are contrasted with the negative, but still informative, results surrounding the influence of average temperature on transmission. Even the best particle from **EIP** produced an EIP curve defined by the parameters c_1 and c_2 that endowed mosquitoes exposed to the disease with immediate infectiousness. In addition, η increased to high values, leading to the model maintaining infection almost exclusively via background human importation. **EIP** also generates evidence for a non-zero amplitude in the transmission function $\beta(t)$, further indicating that

potential variation in the EIP alone does not capture the variation in the original signal.

Despite the failure to explain transmission based solely on temperature-dependent EIP, the heightened impact of background human infection, in both **EIP** and **Control**, suggests an increased importance of importation in general as a transmission dynamics driver. In fact, the isolation of Iquitos promotes the notion that outside infection could indeed have a strong overall impact. Iquitos' main inroad is river travel, which is seasonally affected. Both elevations in river water level (linked to precipitation) and heightened travel occur during the summer months, aligning with the most intense period of dengue transmission. In response, **Control** $_{\lambda_G}$ begins to explore how to explicitly disentangle the internal infectious population from the seasonal external infectious population in the transmission. The best reproduction of the original data signal for a model investigating control measures came from **Control** $_{\lambda_G}$, but in exchange for a small increase in accuracy, the **Control** $_{\lambda_G}$ R_0 becomes unrealistically high. The multiplicative interaction of the two β cosine functions in Equation (15), representing seasonality in both the host and vector probabilities of exposure, obfuscates the overall interpretability of the R_0 measure. The interactions that result from the multiplication of the β functions potentially duplicate the tendencies within the true nature of transmission.

Control $_{\lambda_G, H}$ and **Control** $_{b_0}$ take the investigation of seasonal variation on FOI further by completely eliminating seasonality from transmission, keeping it exclusively on human importation. Although **Control** $_{b_0}$ improved upon **Control** $_{\lambda_G, H}$ by adding another parameter to distinguish between host and vector average transmission values, taking the results of these models together, there is lack of support for seasonality being exclusively due to importation and for removing the seasonality from transmission. Thus, **Control**, with less complexity and assumptions regarding seasonality, is preferable to **Control** $_{b_0}$, despite close AIC scores.

4.2 Modeling Process

The general proximity of scores between all models indicates that each reproduce the original time series behavior to a relatively similar degree. Nevertheless, the goal in this thesis was not to capture the entire data signal exactly but to capture maximal informational meaning at minimal computational expense. As such, there is a delicate balance between the accuracy of reproduction and the quality of information extracted when developing these mechanistic models. We approached this issue by limiting particle parameter selection, from 7 to 9 total, to protect feasible runtimes, around 24 hours running in parallel on an 8 core machine. Embedded in restricting particle length is the simplification of many mechanistic processes within the model because all other non-particle parameters are held constant over time. For instance, the average juvenile vector development duration, τ , is known to vary depending on a multitude of conditions such as diet, larval density, and temperature [48]. Yet, all these subtleties are omitted for convenience. With more computation, mechanistic intuition, and time, the inclusion of more particle parameters could potentially render a model that more accurately reproduces the original signal and yields more substantial conclusions. On the other hand, over-modeling has its risks. On top of the possibility of more computation for less reward, simply adding more parameters may not be productive. A model can fail not because it is ill-posed but because the supporting evidence for a particular phenomenon does not exist in the data. The inherent nature of the data under question limits the potential information that can be garnered and consequently the hypotheses that can be tested.

In evaluating the model results, it is important to remember that the particles selected via the MIF process are merely estimates of the theoretical MLE in the parameter space. The optimal particle selections of the models presented here are rather crude estimates, based on the overall low identifiability of the found parameters. Of course, given enough computational time, MIF will produce results that

converge to the theoretical particle that maximizes the log-likelihood score for the model. Nonetheless, it is important to note that inherent in MIF is an overall struggle toward particle reproducibility, i.e. arriving at the same exact optimal particle in repeated MIF explorations of parameter space. The stochasticity inherent in disease transmission and its modeling compounds the difficulty of finding the singular best particle to resemble the data. In a related fashion, the initial values of the particle swarm can also play a role in the selection of the MLE. Critical to the issue is that within the parameter space there exists both local and global extrema. It is often difficult to know where MIF has chosen the MLE without further investigation of the parameters.

4.3 Future Work

Our work demonstrates the flexibility and power of plug-and-play inference methods. Through these methods, the optimization of model parameter values with interpretable meaning readily test specific hypotheses regarding seasonal dynamics of dengue transmission in Iquitos, Peru. In fact, only by building specific mechanisms into a model like control and human importation of the virus were we able to consider and confirm the impact of these perhaps previously overlooked or undervalued contributing features. However, as noted, the presented model framework only begins to test and uncover ideas behind seasonal dengue transmission in Iquitos for this dataset.

Several improvements and further questions exist and are within reach. First, although we worked from the assumption of small, non-zero initial exposed and infected populations, more investigation into the initial ratio of susceptible to exposed to infected humans could supply better results. The initial distribution of human population classes is important because it dictates the intensity of disease, and hence overall shape, of the output at the onset of the signal. Currently, a non-identifiable S_0 encourages wide value ranges in the rest of the particle, particularly η . A better

Chapter 4. Discussion

assignment of initial populations could contract the scope of parameter values and even refine the overall log-likelihood of the models.

Second, the strong b_1 signal indicates that more detail in the transmission curve β could benefit model quality in the pursuit of seasonal transmission drivers. While the single cosine function in the current models focuses on intra-annual variation, building in an additional, longer period cosine signal might more successfully capture previously omitted inter-annual behavior.

Third, neglected in these models are the effects of cross-immunity in humans between serotypes. Cross-immunity plays an important role in a more detailed understanding of dengue transmission. Adding cross-immunity to the model would further inform the initial population classes and could readily be applied through an added partial recovery class and a human age structure system. For instance, in Thailand, and similarly with many other endemic regions, dengue fever is so prevalent that most of the population has a first infection before age 15 [49]. As a result, there is a strong effect of built-in immunity throughout the population that could play a role in transmission, especially with the rotating serotype dominance pattern seen in Iquitos. Moreover, another study on this same incidence data in Iquitos shows that differences between serotypes also affect treatment strategies in control and vaccination [50].

Finally, as an ultimate objective of mathematical model research in epidemiology is to strengthen the response to public health risks through an enlightened reaction to disease, further hypotheses using the control mechanism feature can offer suggestions as to the optimal timing and strength of control to reduce dengue cost and morbidity.

Figures

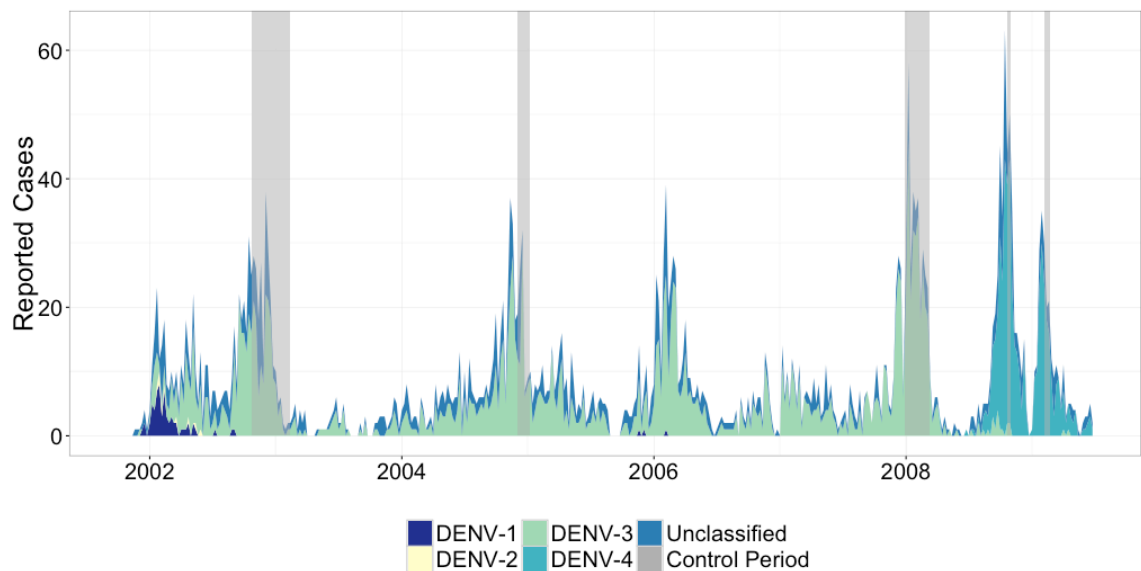


Figure 1: Case reports in Iquitos, Peru from 1 October 2001 to 1 July 2009 by serotype. Each serotype time series is stacked to show the breakdown of total incidence by type over time. The domination of DENV-3 starts in the summer months of 2002 and continues for several seasons, eventually replaced by DENV-4. Seasonal periods are irregular with most peaks in incidence occurring in the warmer summer months. Control fumigation (grey bands), consisted of household insecticide spraying on a citywide scale. There is clear pattern of reduced incidence after each control period.

Figures

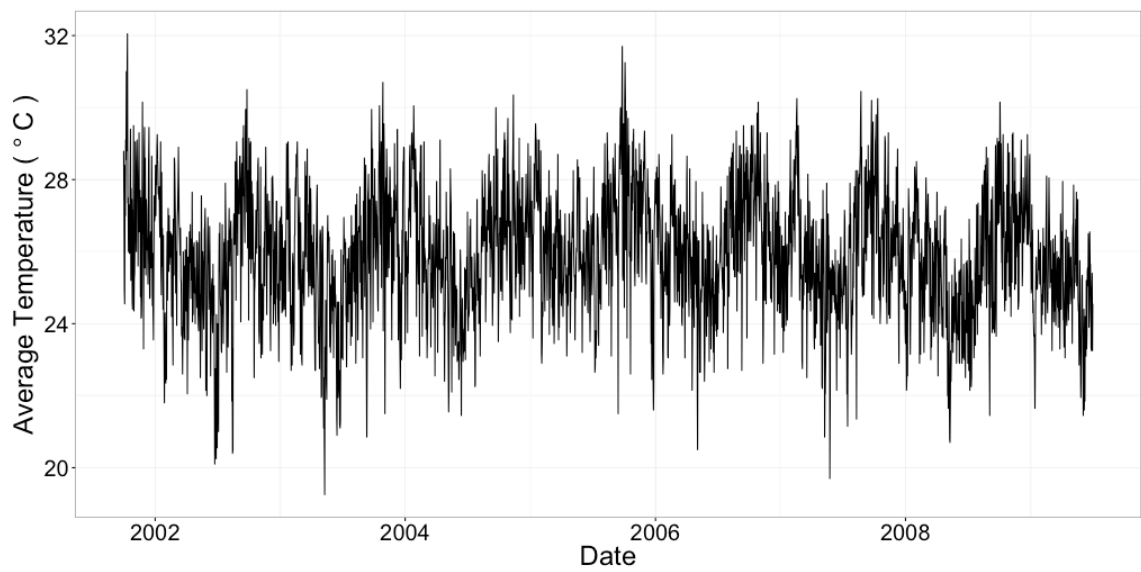
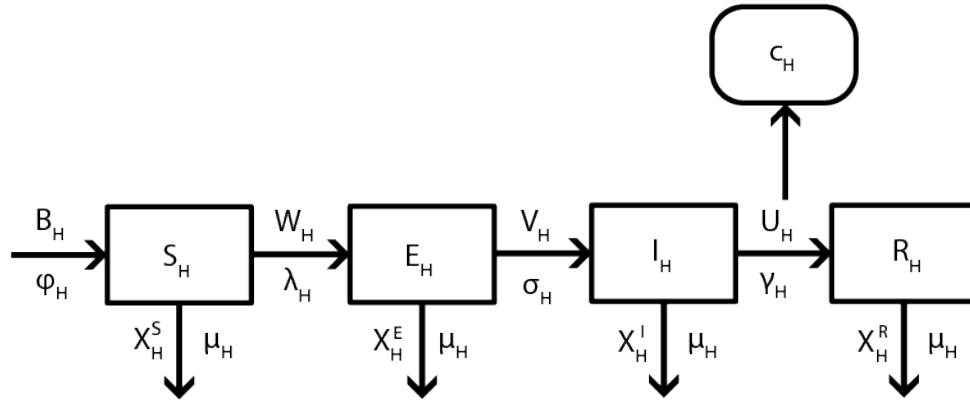
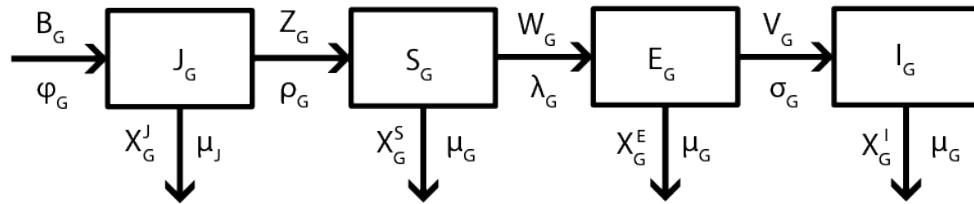


Figure 2: Average temperature ($^{\circ}\text{C}$) in Iquitos, Peru from 1 October 2001 to 1 July 2009. Average temperature is calculated as the mean of the daily low and high temperature recorded at the Iquitos airport US National Oceanic and Atmospheric Administration (NOAA) weather station. The limited variation in average temperature ranges largely between 20 to 30 $^{\circ}\text{C}$ (68 to 86 $^{\circ}\text{F}$).

Figures



(a) Human SEIR model.



(b) Mosquito JSEI model.

Figure 3: Demographic host-vector model. The compartments show flow between states of infection for (a) the human population and (b) the vector population. The arrows indicate directionality of flow. Descriptions above or to the left of each arrow signify the population that moves between compartments, and the description below or to the right of the arrows signify the probabilities of transition. See §A.1 for a summary of the model equations.

Figures

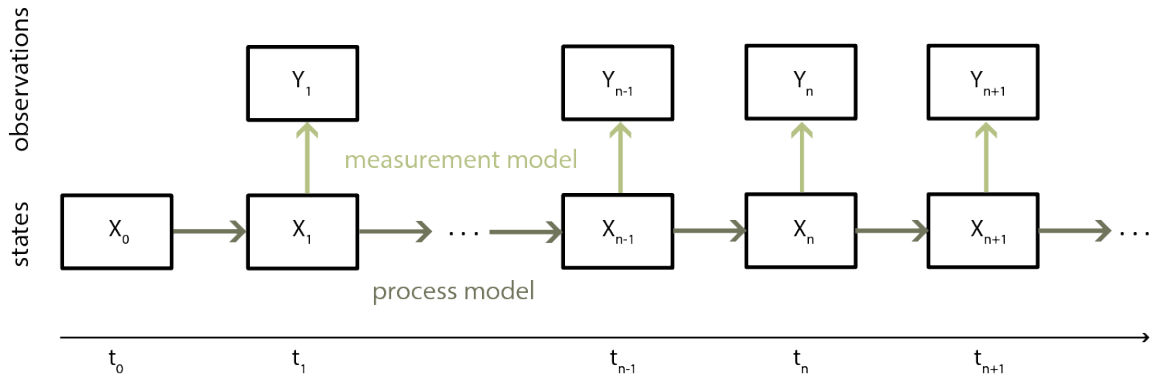


Figure 4: Schematic diagram of POMDP model. The arrows show variable dependence (a box depends on the boxes pointing to it). At time t_n , measurements, Y_n , depend on the state, X_n . The distribution of the measurement Y_n is conditional on X_n and independent of all other variables. As a Markov chain the distribution of the state X_{n+1} is conditional on X_n and independent of the values of $X_k, k < n$ and $Y_k, k \leq n$. The process model (dark green) stipulates movement from state to state, whereas the measurement model (light green) supplies the transition rules from state to observation.

Figures

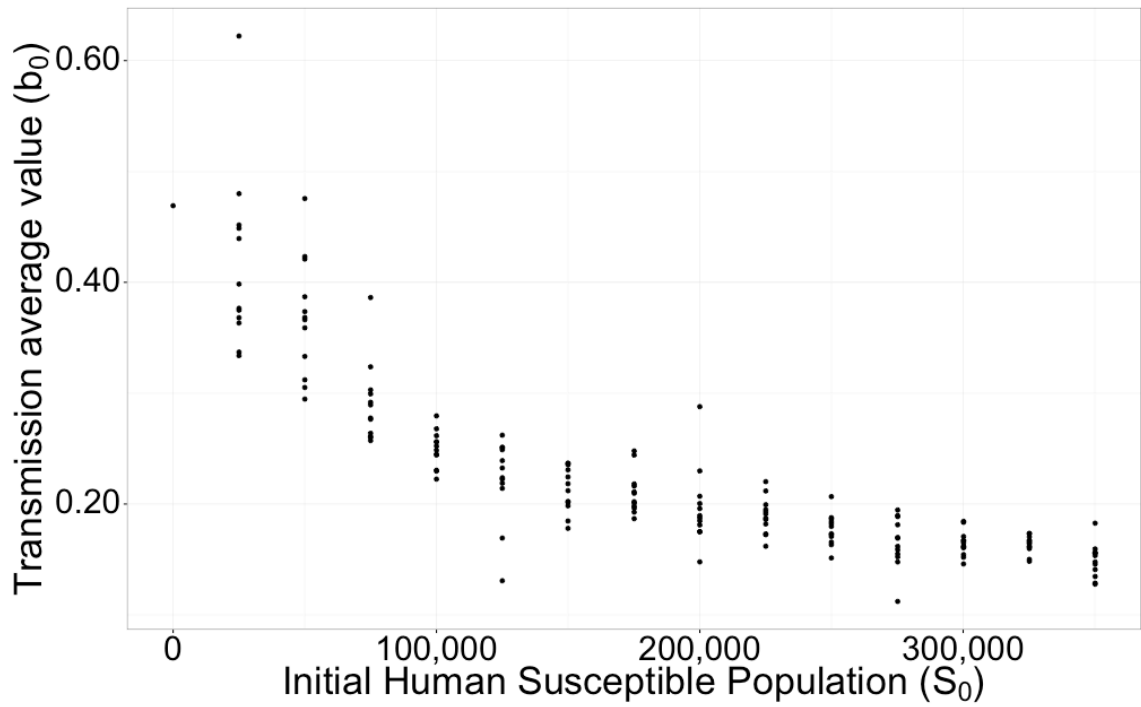


Figure 5: Initial susceptible human population and average seasonal transmission value. $S_H(t_0) = S_0$ and b_0 show a clear pattern of dependence where smaller values of S_0 align with larger values of b_0 .

Figures

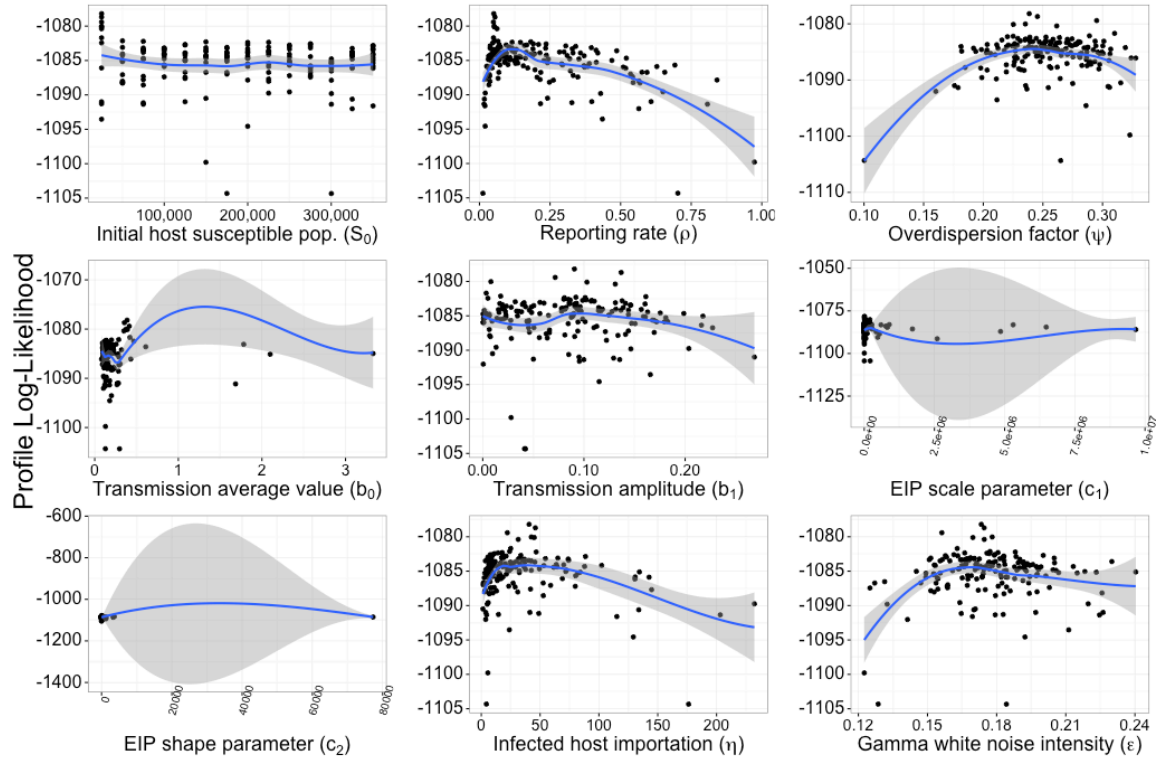
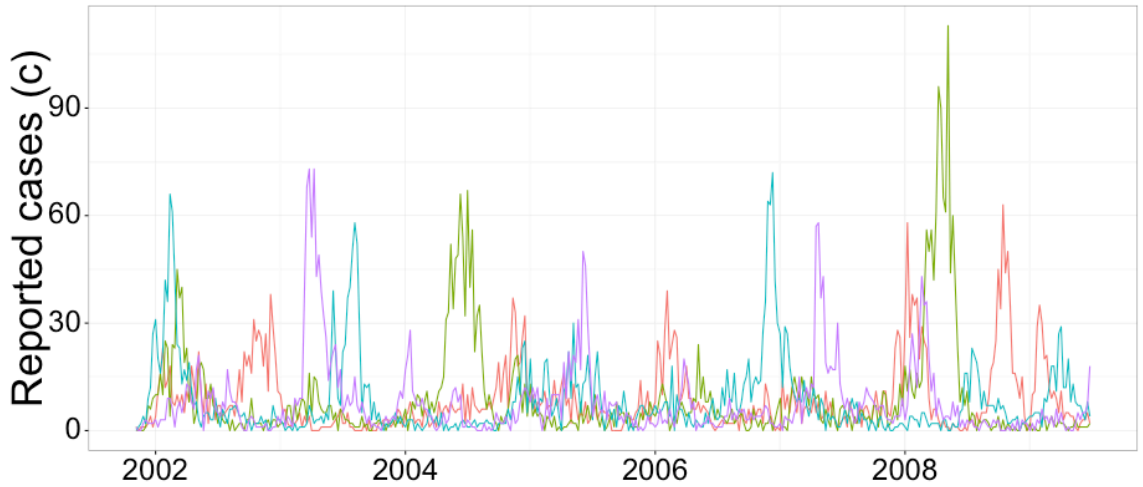
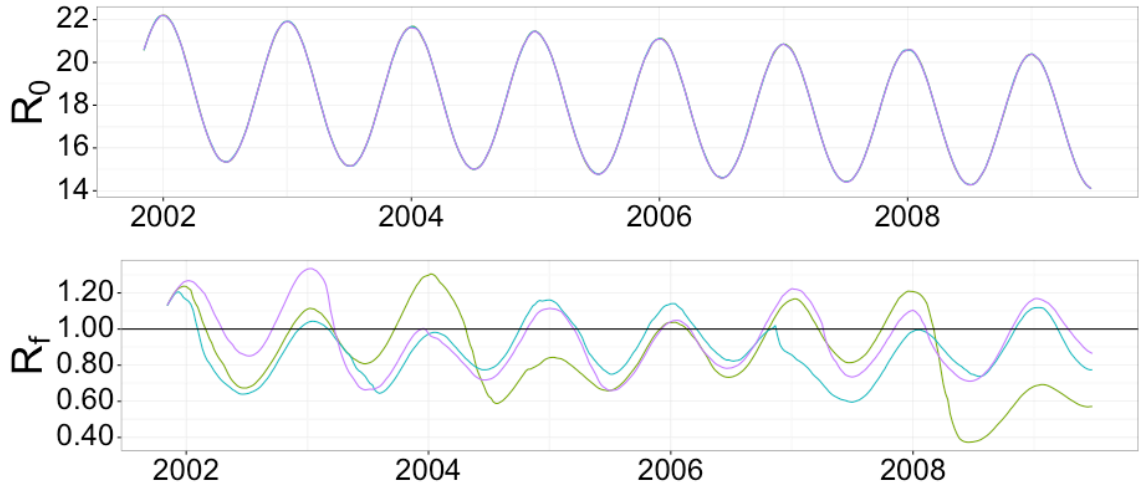


Figure 6: EIP likelihood profiles. The profile is over S_0 with 225 initial particles of $\theta = [S_0, \rho, \psi, b_0, b_1, c_1, c_2, \eta, \epsilon]$.

Figures



(a) **EIP** case report simulated output. Three simulations of case reports shown (purple, cyan, olive) over time compared to the original data (red).



(b) Simulations of temperature-dependent **EIP** R_0 and R_f . R_0 is much larger than expected for dengue, 3 – 5. The black line at 1 in the R_f plot illustrates when the infectious population of the model is increasing (above the line) or decreasing (below the line).

Figure 7: Three simulations of the **EIP**: (a) reported cases, (b) reproductive numbers, R_0 . and R_f .

Figures

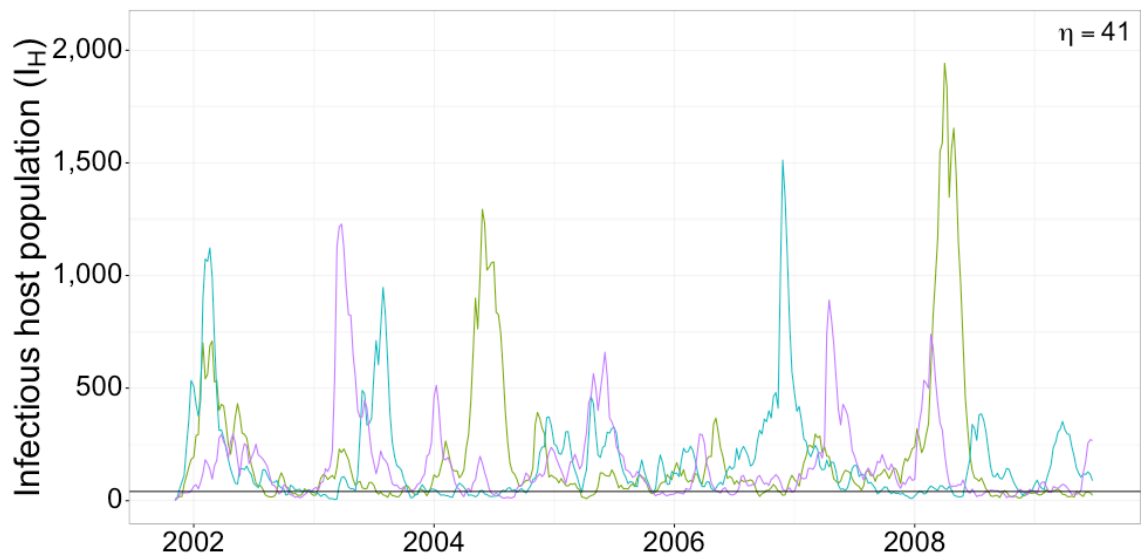


Figure 8: EIP model background infection. Three example simulations of infected human population counts (purple, cyan, olive) over time compared to the constant human importation parameter, in this case $\eta \approx 41$, assigned by the model from the maximal point estimate particle. Infection continually renews and rarely dips beneath the background infection level.

Figures

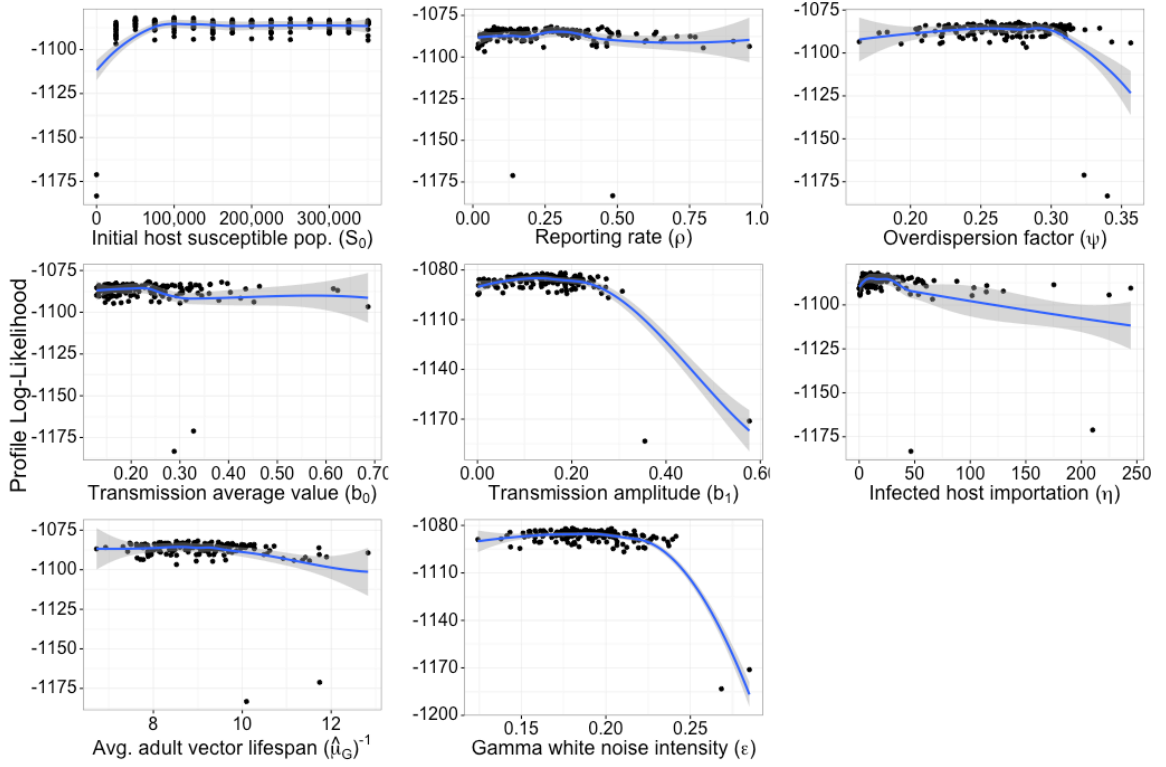
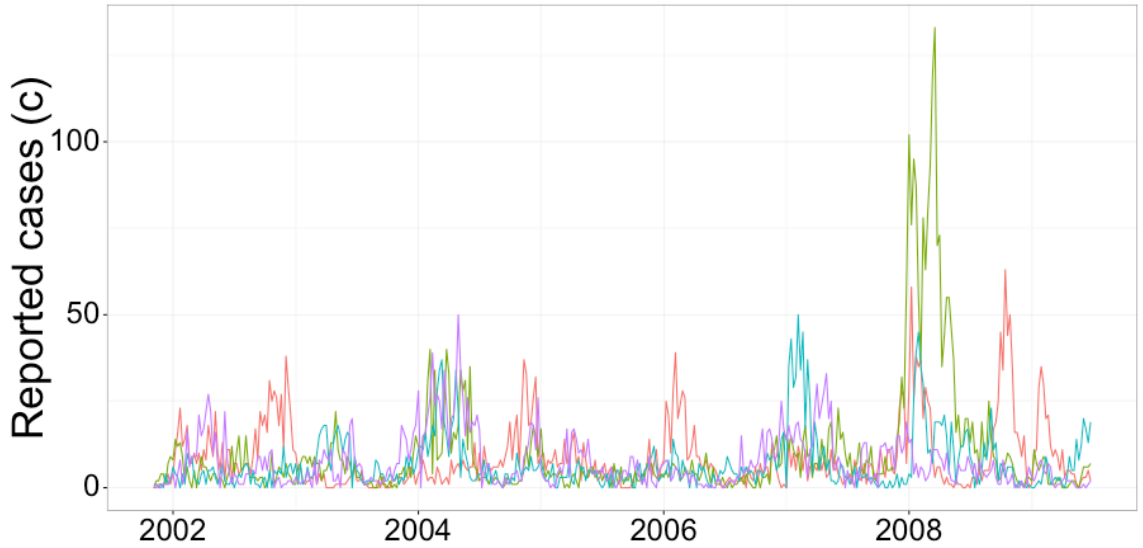
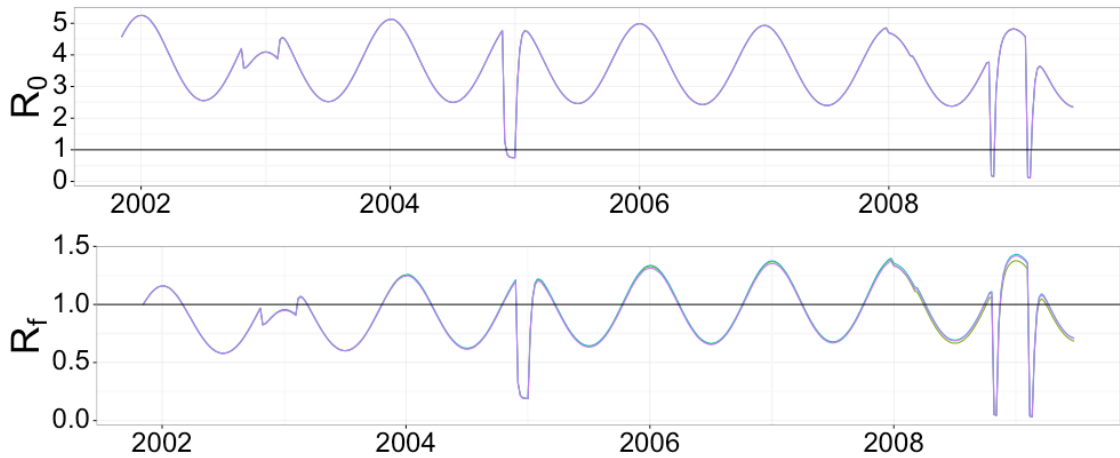


Figure 9: Control likelihood profiles. The profile is over S_0 with 225 initial particles of $\theta = [S_0, \rho, \psi, b_0, b_1, \eta, (\hat{\mu}_G)^{-1}, \epsilon]$.

Figures



(a) **Control** case report simulated output. Three simulations of case reports shown (purple, cyan, olive) over time compared to the original data (red).



(b) Simulation of **Control** model R_0 and R_f . The control measures are visible in both curves. The black line at 1 in the R_0 plot illustrates when, at the given rate, the infection will die out (below the line) or spread (above the line) in the population over the long-term.

Figure 10: Three simulations of the **Control**: (a) reported cases, (b) reproductive numbers, R_0 and R_f .

Figures

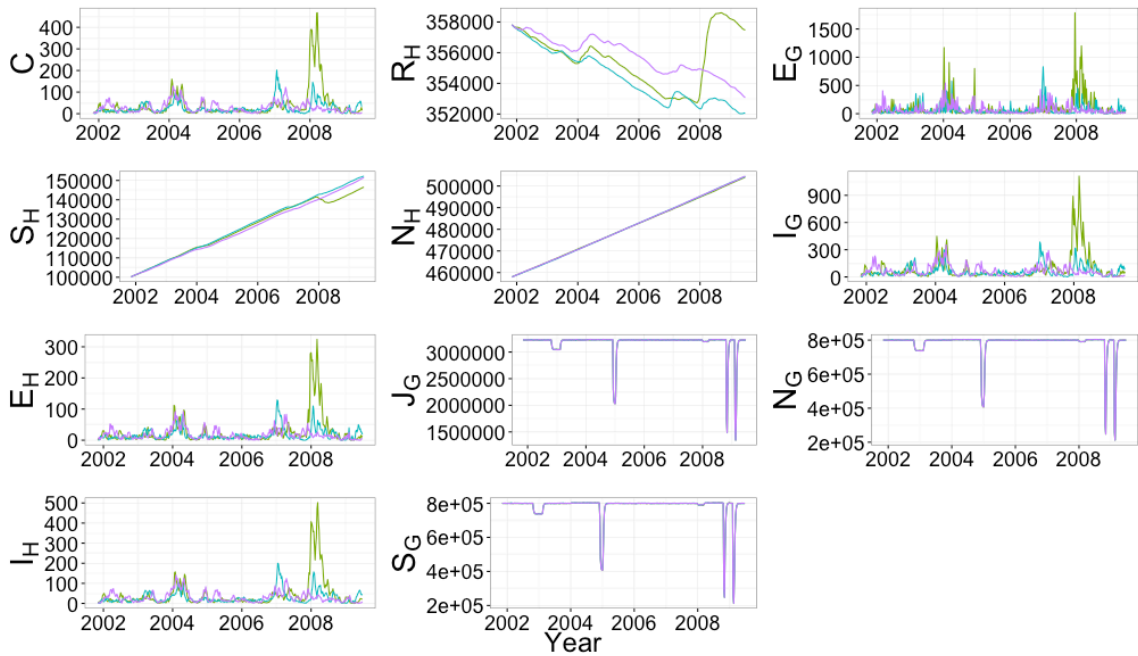


Figure 11: Control states simulations. Three simulations (purple, cyan, olive) of **Control** states, including total incidence (C), states of infection for the mosquito population (J_G , S_G , E_G , I_G , and N_G), and states of infection for the human population (S_H , E_H , I_H , R_H , and N_H). The control signal is visible in the mosquito populations J_G , S_G , and N_G .

Figures

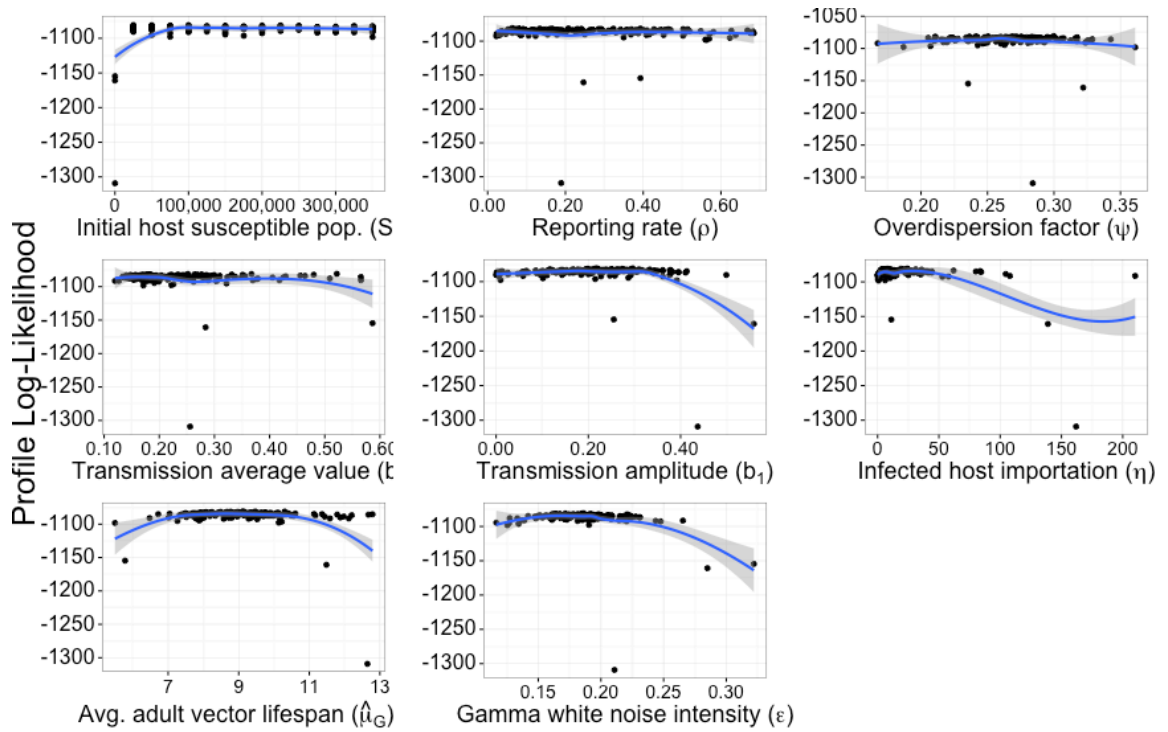
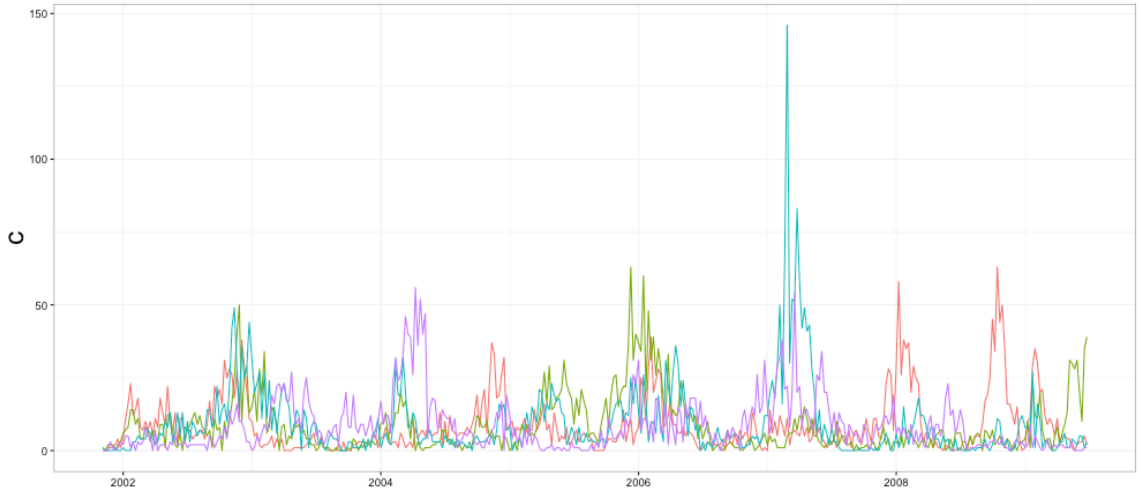
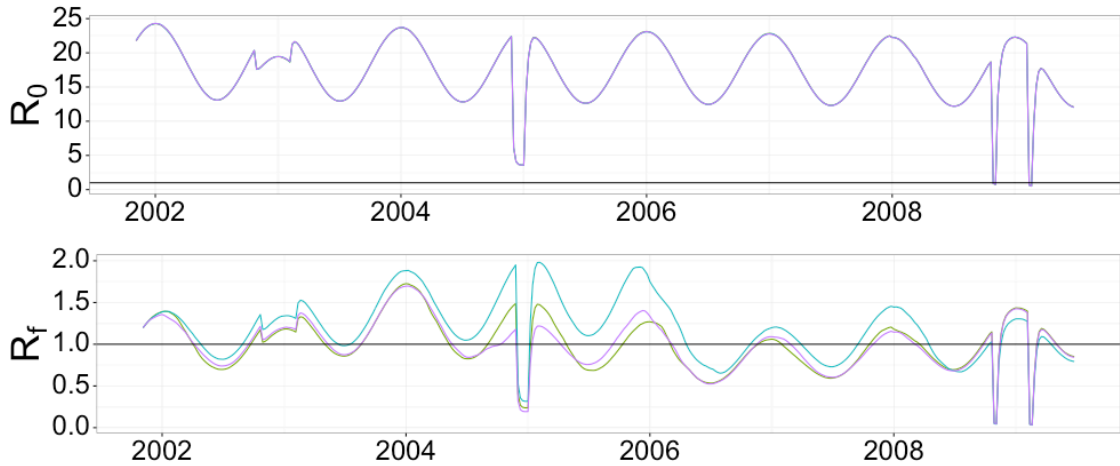


Figure 12: Control $_{\lambda_G}$ likelihood profiles. The profile is over S_0 with 225 initial particles of $\theta = [S_0, \rho, \psi, b_0, b_1, \eta, (\hat{\mu}_G)^{-1}, \epsilon]$.

Figures



(a) **Control** λ_G case report simulated output. Three simulations of case reports shown (purple, cyan, olive) over time compared to the original data (red).



(b) Simulations of **Control** λ_G case reports, R_0 and R_f . The level of R_0 noticeably increases over the R_0 from **Control**. A higher than expected, usually 3 – 5, indicates a lack of parameter identifiability for this model.

Figure 13: Three simulations of the **Control** λ_G : (a) reported cases, (b) reproductive numbers, R_0 . and R_f .

4.4 Control $_{\lambda_{G,H}}$ Plots

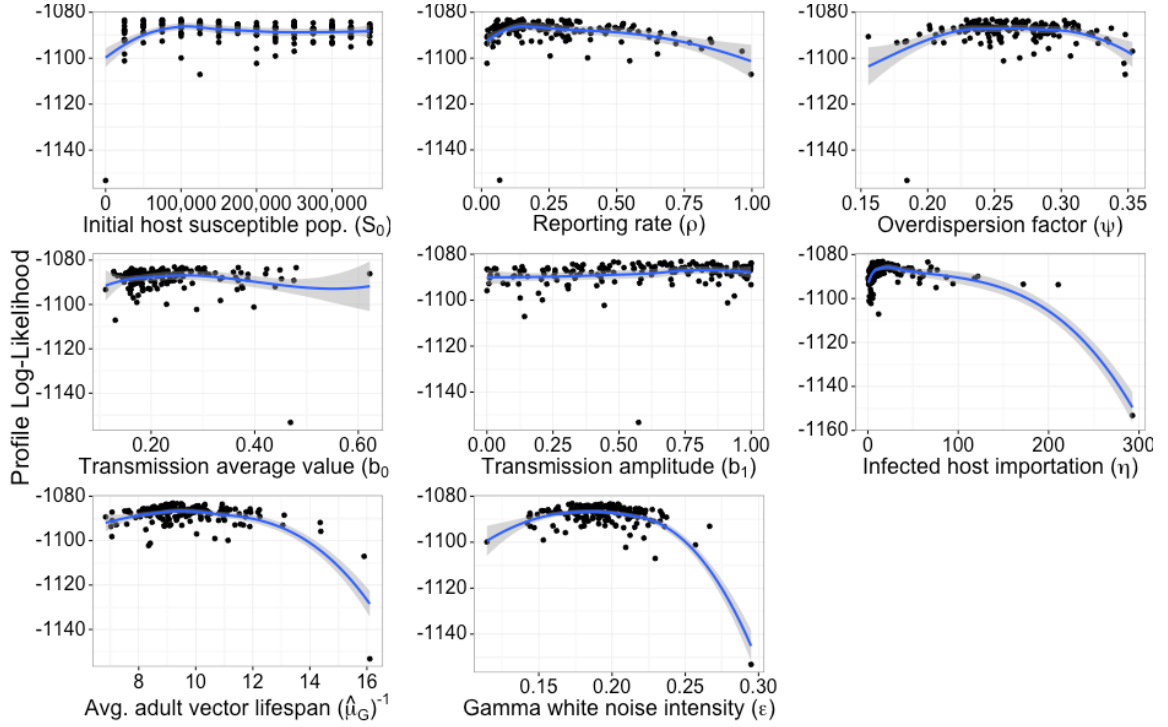
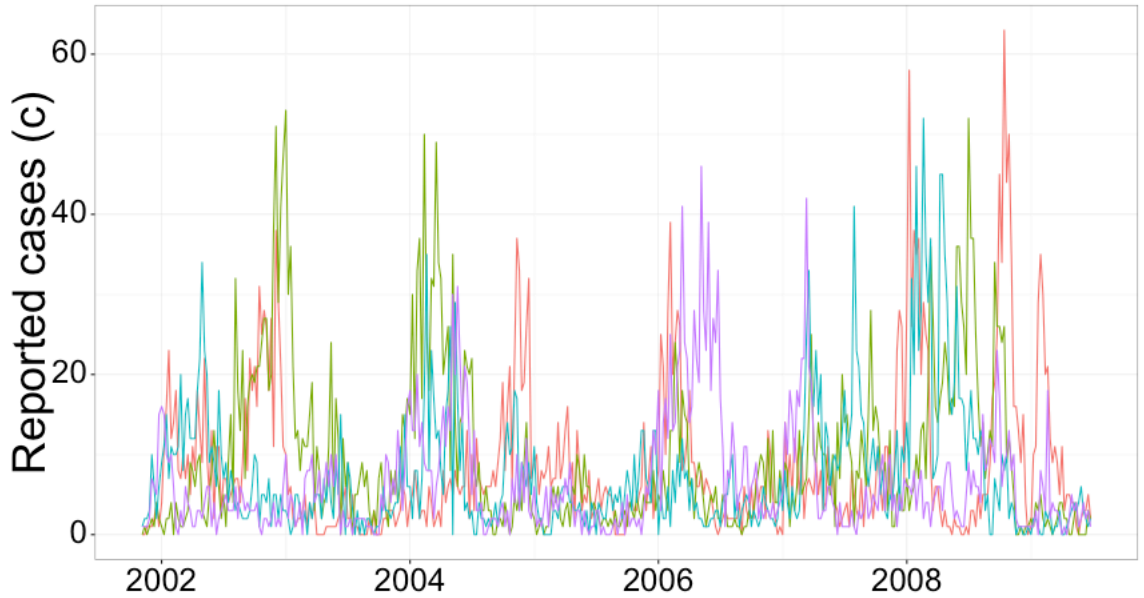
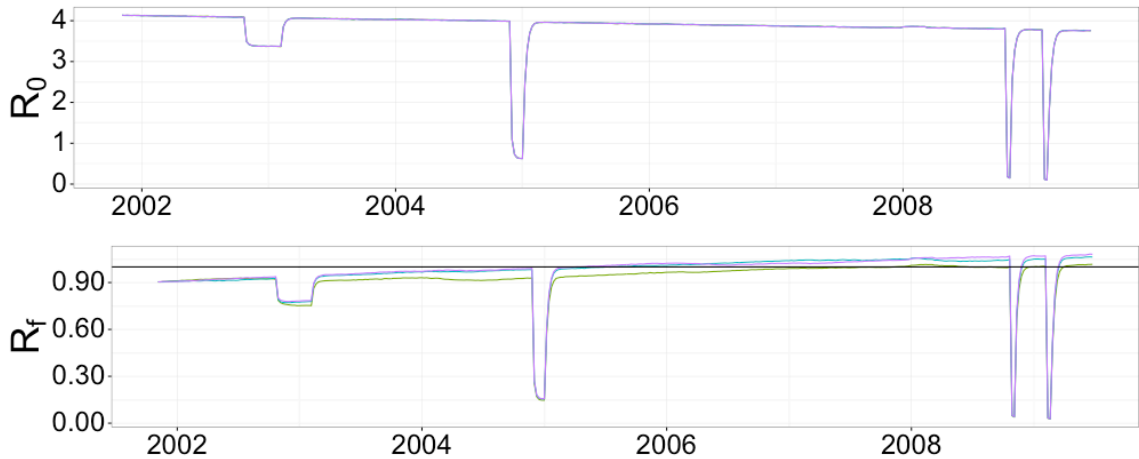


Figure 14: Control $_{\lambda_{G,H}}$ likelihood profiles. The profile is over S_0 with 225 initial particles of $\theta = [S_0, \rho, \psi, b_0, b_1, \eta, (\hat{\mu}_G)^{-1}, \epsilon]$.

Figures



(a) **Control** $_{\lambda_{G,H}}$ case report simulated output. Three simulations of case reports shown (purple, cyan, olive) over time compared to the original data (red).



(b) Simulations of **Control** $_{\lambda_{G,H}}$ case reports, R_0 and R_f . The removal of seasonality in transmission is also visibly removed from the reproductive numbers.

Figure 15: Three simulations of the **Control** $_{\lambda_{G,H}}$: (a) reported cases, (b) reproductive numbers, R_0 . and R_f .

4.5 Control $_{b_0}$ Plots

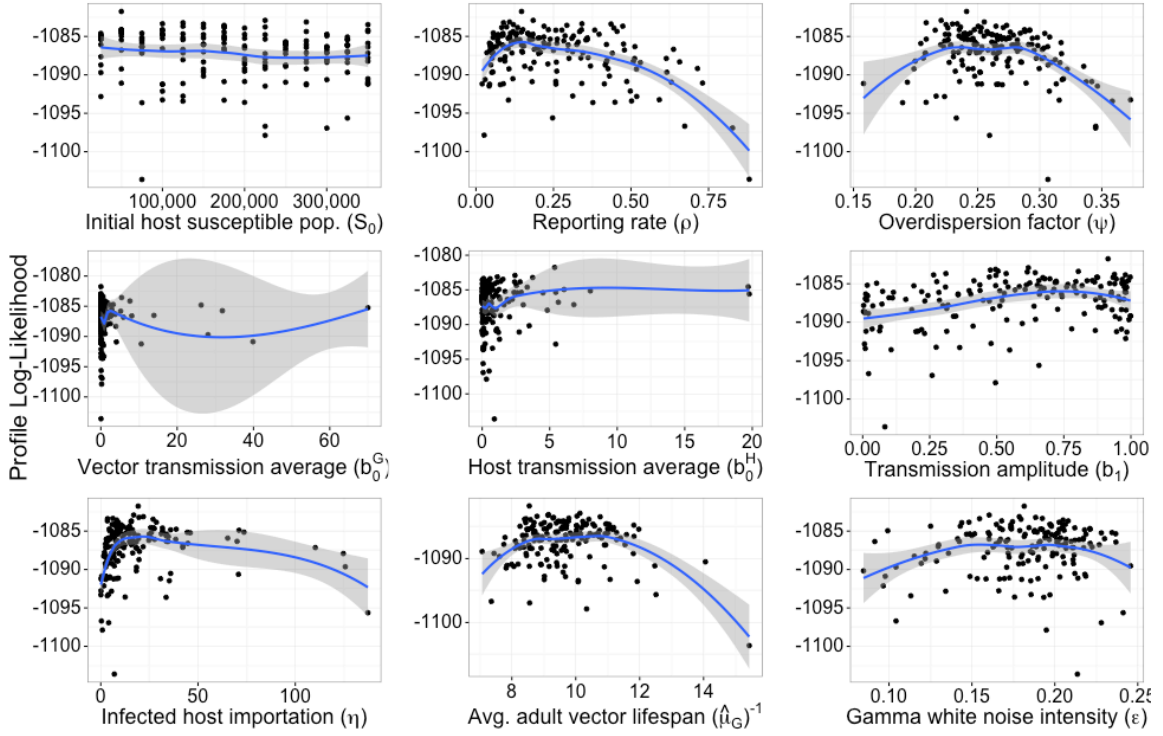
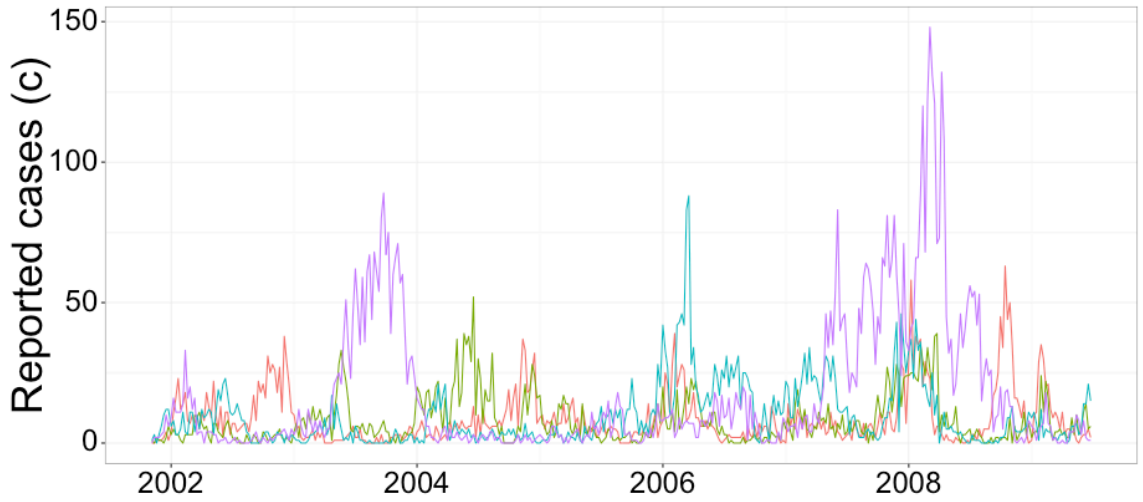
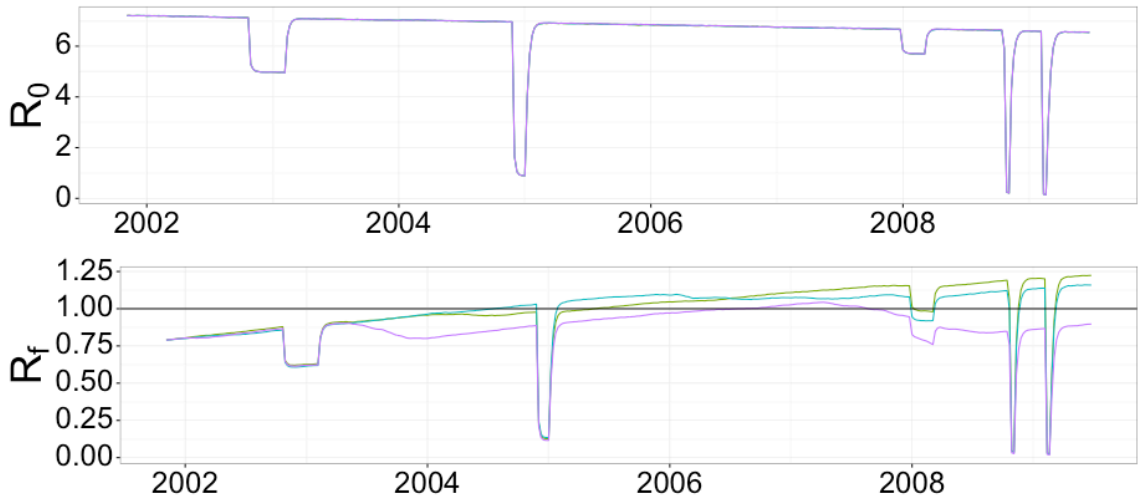


Figure 16: Control $_{b_0}$ likelihood profiles. The profile is over S_0 with 225 initial particles of $\theta = [S_0, \rho, \psi, b_0^H, b_0^G, b_1, \eta, (\hat{\mu}_G)^{-1}, \epsilon]$.

Figures



(a) **Control**_{b₀} case report simulated output. Three simulations of case reports shown (purple, cyan, olive) over time compared to the original data (red).



(b) Simulations of **Control**_{b₀} case reports, R_0 and R_f .

Figure 17: Three simulations of the **Control**_{b₀}: (a) reported cases, (b) reproductive numbers, R_0 . and R_f .

Tables

Table 1: Summary of citywide fumigation efforts

Period, i	Dates (Total Days)	Household effort	Weight, w_i
1	10/23/2002 – 02/10/2003 (111)	55,743	1
2	12/01/2004 – 01/05/2005 (36)	35,572	1.9676
3	12/27/2007 – 03/08/2008 (73)	33,633	0.91744
4	10/20/2008 – 11/03/2008 (15)	24,816	3.2944
5	02/05/2009 – 02/21/2009 (17)	32,350	3.7893

Table 2: Parameter list for demographic model

Parameter	Description	Value (Units)	Source
t	Time	1 (days)	—
$T(t)$	Daily avg. temperature	(C°)	—
$\hat{\phi}_H$	Human birth rate, Peru (2014)	5.0877×10^{-5} (days ⁻¹)(person ⁻¹)	[51]
$\hat{\mu}_H$	Human death rate, Peru (2014)	1.6411×10^{-5} (days ⁻¹)(person ⁻¹)	[51]
$(\hat{\sigma}_H)^{-1}$	Avg. intrinsic incubation period	4 (days)	[52]
$(\hat{\gamma}_H)^{-1}$	Avg. duration of infectiousness	7 (days)	[53, 54]
$\hat{\phi}_G$	Avg. vector egg production rate	4.6 (days ⁻¹)(mosquito ⁻¹)	[55]
α	Juvenile vector competition	4.2534×10^{-7} (days ⁻¹)(larva ⁻¹)	§A.3
$\hat{\theta}_G$	Avg. egg survivorship	0.68	[56]
$(\hat{\mu}_J)^{-1}$	Avg. juvenile vector lifespan	45 (days)	§A.2
$(\tau)^{-1}$	Avg. juvenile vector development	10 (days)	[48]
$(\hat{\mu}_G)^{-1}$	Avg. adult vector lifespan	10 (days)	[57, 58]
$(\hat{\sigma}_G)^{-1}$	Avg. extrinsic incubation period	9 (days)	[59]

Table 3: Altered **Control** Force of Infection hypotheses summary

Model	F_H	F_G
EIP	$\beta(t) \frac{I_G}{N_H}$	$\beta(t)(I_H + \eta) \frac{\zeta(\epsilon)}{N_H}$
Control		
Control λ_G	$\beta(t) \frac{I_G}{N_H}$	$(b_0 I_H + \beta(t)\eta) \frac{\zeta(\epsilon)}{N_H}$
Control $\lambda_{G,H}$	$b_0 \frac{I_G}{N_H}$	$(b_0 I_H + \beta(t)\eta) \frac{\zeta(\epsilon)}{N_H}$
Control b_0	$b_0^H \frac{I_G}{N_H}$	$(b_0^G I_H + \beta(t)\eta) \frac{\zeta(\epsilon)}{N_H}$

Table 4: Maximum point estimates for model parameters

Model	S_0	ρ	ψ	b_0	b_0^H	b_0^G	b_1	c_1	c_2	η	$(\hat{\mu}_G)^{-1}$	ϵ
EIP	25,010	0.051	0.24	0.39	—	—	0.09	8140.91	4.27	40.61	—	0.17
Control	100,008	0.27	0.25	0.23	—	—	0.18	—	—	10.02	9.00	0.18
Control λ_G	25,010	0.063	0.30	0.52	—	—	0.30	—	—	24.33	9.11	0.19
Control $\lambda_{G,H}$	100,008	0.15	0.25	0.24	—	—	0.73	—	—	16.50	9.20	0.17
Control b_0	50,012	0.15	0.24	—	0.019	5.36	0.91	—	—	19.23	8.55	0.18

Table 5: Model comparisons: profile log-likelihood and AIC scores

Model	$\hat{\ell}(\hat{\theta})$	AIC
EIP	-1078.19	2174.38
Control	-1081.80	2179.60
Control $_{\lambda_G}$	-1080.84	2177.68
Control $_{\lambda_{G,H}}$	-1083.10	2182.20
Control $_{b_0}$	-1081.75	2181.50

Bibliography

- [1] Duane J Gubler. “Resurgent vector-borne diseases as a global health problem.” In: *Emerging infectious diseases* 4.3 (1998), p. 442.
- [2] Paul E Parham et al. “Climate, environmental and socio-economic change: weighing up the balance in vector-borne disease transmission.” In: *Phil. Trans. R. Soc. B* 370.1665 (2015), p. 20130551.
- [3] Giovanni Rezza. “Dengue and chikungunya: long-distance spread and outbreaks in naive areas.” In: *Pathogens and global health* 108.8 (2014), pp. 349–355.
- [4] Morgan Hennessey, Marc Fischer, and J Erin Staples. “Zika virus spreads to new areas—region of the Americas, May 2015–January 2016.” In: *American Journal of Transplantation* 16.3 (2016), pp. 1031–1034.
- [5] World Health Organization et al. *Dengue: guidelines for diagnosis, treatment, prevention and control*. World Health Organization, 2009.
- [6] Alfonso Guzman and Raul E Istriz. “Update on the global spread of dengue.” In: *International journal of antimicrobial agents* 36 (2010), S40–S42.
- [7] Duane J Gubler. “Dengue and dengue hemorrhagic fever.” In: *Clinical microbiology reviews* 11.3 (1998), pp. 480–496.
- [8] Donald S Shepard et al. “Economic impact of dengue illness in the Americas.” In: *The American journal of tropical medicine and hygiene* 84.2 (2011), pp. 200–207.
- [9] Jose A Suaya, Donald S Shepard, and Mark E Beatty. “Dengue: burden of disease and costs of illness.” In: *TDR. Report of the Scientific Working Group Meeting on Dengue*. 2007, pp. 35–49.

BIBLIOGRAPHY

- [10] Steven T Stoddard et al. “Long-term and seasonal dynamics of dengue in Iquitos, Peru.” In: *PLoS Negl Trop Dis* 8.7 (2014), e3003.
- [11] Oliver J Brady et al. “Refining the global spatial limits of dengue virus transmission by evidence-based consensus.” In: *PLoS Negl Trop Dis* 6.8 (2012), e1760.
- [12] Louis Lambrechts, Thomas W Scott, and Duane J Gubler. “Consequences of the expanding global distribution of *Aedes albopictus* for dengue virus transmission.” In: *PLoS Negl Trop Dis* 4.5 (2010), e646.
- [13] Robert V Gibbons et al. “Analysis of repeat hospital admissions for dengue to estimate the frequency of third or fourth dengue infections resulting in admissions and dengue hemorrhagic fever, and serotype sequences.” In: *The American journal of tropical medicine and hygiene* 77.5 (2007), pp. 910–913.
- [14] Nicholas G Reich et al. “Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity.” In: *Journal of The Royal Society Interface* 10.86 (2013), p. 20130414.
- [15] World Health Organization. “Dengue vaccine: WHO position paper–July 2016.” In: *Weekly epidemiological record* 30.91 (July 2016), pp. 349–364. URL: <http://www.who.int/wer/2016/wer9130.pdf?ua=1>.
- [16] Sai Kit Lam. “Challenges in reducing dengue burden; diagnostics, control measures and vaccines.” In: *Expert review of vaccines* 12.9 (2013), pp. 995–1010.
- [17] Maria G Guzman et al. “Dengue: a continuing global threat.” In: *Nature Reviews Microbiology* 8 (2010), S7–S16.
- [18] Michael A Johansson, Derek AT Cummings, and Gregory E Glass. “Multi-year climate variability and dengue–El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis.” In: *PLoS Med* 6.11 (2009), e1000168.
- [19] Karen M Campbell et al. “The complex relationship between weather and dengue virus transmission in Thailand.” In: *The American journal of tropical medicine and hygiene* 89.6 (2013), pp. 1066–1080.

BIBLIOGRAPHY

- [20] Michael A Johansson, Francesca Dominici, and Gregory E Glass. “Local and global effects of climate on dengue transmission in Puerto Rico.” In: *PLoS Negl Trop Dis* 3.2 (2009), e382.
- [21] S Christophers et al. “*Aedes aegypti* (L.) the yellow fever mosquito: its life history, bionomics and structure.” In: *Rickard*. (1960).
- [22] Micha Bar-Zeev. “The effect of temperature on the growth rate and survival of the immature stages of *Aedes aegypti* (L.)” In: *Bulletin of Entomological Research* 49.01 (1958), pp. 157–163.
- [23] LM Rueda et al. “Temperature-dependent development and survival rates of *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae).” In: *Journal of medical entomology* 27.5 (1990), pp. 892–898.
- [24] W Tun-Lin, TR Burkot, and BH Kay. “Effects of temperature and larval diet on development rates and survival of the dengue vector *Aedes aegypti* in north Queensland, Australia.” In: *Medical and veterinary entomology* 14.1 (2000), pp. 31–37.
- [25] RK Walsh et al. “Assessing the impact of density dependence in field populations of *Aedes aegypti*.” In: *Journal of Vector Ecology* 36.2 (2011), pp. 300–307.
- [26] A Rohani et al. “The effect of extrinsic incubation temperature on development of dengue serotype 2 and 4 viruses in *Aedes aegypti* (L.)” In: *Southeast Asian Journal of Tropical Medicine and Public Health* 40.5 (2009), p. 942.
- [27] Pedro Barbosa, T Michael Peters, and NC Greenough. “Overcrowding of mosquito populations: responses of larval *Aedes aegypti* to stress.” In: *Environmental Entomology* 1.1 (1972), pp. 89–93.
- [28] Roy M Anderson, Robert M May, and B Anderson. *Infectious diseases of humans: dynamics and control*. Vol. 28. Wiley Online Library, 1992.
- [29] Michael A Robert et al. “Modeling Mosquito-Borne Disease Spread in US Urbanized Areas: The Case of Dengue in Miami.” In: *PloS one* 11.8 (2016), e0161365.

BIBLIOGRAPHY

- [30] Ray Hilborn and Marc Mangel. *The ecological detective: confronting models with data*. Vol. 28. Princeton University Press, 1997.
- [31] Carles Bret et al. “Time series analysis via mechanistic models.” In: *The Annals of Applied Statistics* (2009), pp. 319–348.
- [32] Tim Coulson, Pejman Rohani, and Mercedes Pascual. “Skeletons, noise and population growth: the end of an old debate?” In: *Trends in Ecology & Evolution* 19.7 (2004), pp. 359–364.
- [33] Daihai He, Edward L Ionides, and Aaron A King. “Plug-and-play inference for disease dynamics: measles in large and small populations as a case study.” In: *Journal of the Royal Society Interface* (2009).
- [34] Gavin J Gibson and Eric Renshaw. “Likelihood estimation for stochastic compartmental models using Markov chain methods.” In: *Statistics and Computing* 11.4 (2001), pp. 347–358.
- [35] Katia Koelle and Mercedes Pascual. “Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera.” In: *The American Naturalist* 163.6 (2004), pp. 901–913.
- [36] Alexandros Beskos et al. “Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion).” In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.3 (2006), pp. 333–382.
- [37] Aaron A King. *Model-based Inference in Ecology and Epidemiology*. 2016 (accessed September 1, 2016). URL: <http://kingaa.github.io/short-course/>.
- [38] EL Ionides, C Bret, and AA King. “Inference for nonlinear dynamical systems.” In: *Proceedings of the National Academy of Sciences* 103.49 (2006), pp. 18438–18443.
- [39] Aaron A King, Dao Nguyen, and Edward L Ionides. “Statistical inference for partially observed Markov processes via the R package pomp.” In: *arXiv preprint arXiv:1509.00503* (2015).

BIBLIOGRAPHY

- [40] FMG Magpantay et al. “Pertussis immunity and epidemiology: mode and duration of vaccine-induced immunity.” In: *Parasitology* 143.07 (2016), pp. 835–849.
- [41] Micaela Martinez-Bakker, Aaron A King, and Pejman Rohani. “Unraveling the transmission ecology of polio.” In: *PLoS Biol* 13.6 (2015), e1002172.
- [42] Sourya Shrestha et al. “The role of influenza in the epidemiology of pneumonia.” In: *Scientific reports* 5 (2015).
- [43] Aaron A King et al. “Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola.” In: *Proc. R. Soc. B*. Vol. 282. 1806. The Royal Society. 2015, p. 20150347.
- [44] Brett M Forshey et al. “Arboviral etiologies of acute febrile illnesses in Western South America, 2000–2007.” In: *PLoS Negl Trop Dis* 4.8 (2010), e787.
- [45] Irving Phillips et al. “First documented outbreak of dengue in the Peruvian Amazon region.” In: (1992).
- [46] Helen Abbey. “An examination of the Reed-Frost theory of epidemics.” In: *Human biology* 24.3 (1952), p. 201.
- [47] Edward L Ionides et al. “Inference for dynamic and latent variable models via iterated, perturbed Bayes maps.” In: *Proceedings of the National Academy of Sciences* 112.3 (2015), pp. 719–724.
- [48] Jannelle Couret, Ellen Dotson, and Mark Q Benedict. “Temperature, larval diet, and density effects on development rate and survival of *Aedes aegypti* (Diptera: Culicidae).” In: *PLoS One* 9.2 (2014), e87468.
- [49] P Pongsumpun and IM Tang. “Transmission of dengue hemorrhagic fever in an age structured population.” In: *Mathematical and Computer Modelling* 37.9 (2003), pp. 949–961.
- [50] Robert C Reiner et al. “Time-varying, serotype-specific force of infection of dengue virus.” In: *Proceedings of the National Academy of Sciences* 111.26 (2014), E2694–E2702.

BIBLIOGRAPHY

- [51] *The World Factbook: PERU*. 2016 (accessed September 1, 2015). URL: <https://www.cia.gov/library/publications/the-world-factbook/geos/pe.html>.
- [52] Kara E Rudolph et al. “Incubation periods of mosquito-borne viral infections: a systematic review.” In: *The American journal of tropical medicine and hygiene* 90.5 (2014), pp. 882–891.
- [53] DJ Gubler et al. “Viraemia in patients with naturally acquired dengue infection.” In: *Bulletin of the World Health Organization* 59.4 (1981), p. 623.
- [54] David W Vaughn et al. “Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity.” In: *Journal of Infectious Diseases* 181.1 (2000), pp. 2–9.
- [55] Linda M Styer et al. “Mortality and reproductive dynamics of *Aedes aegypti* (Diptera: Culicidae) fed human blood.” In: *Vector-borne and zoonotic diseases* 7.1 (2007), pp. 86–98.
- [56] Leon E Hugo et al. “Adult survivorship of the dengue mosquito *Aedes aegypti* varies seasonally in central Vietnam.” In: *PLoS Negl Trop Dis* 8.2 (2014), e2669.
- [57] Lynda E Muir and Brian H Kay. “*Aedes aegypti* survival and dispersal estimated by mark-release-recapture in northern Australia.” In: *The American Journal of Tropical Medicine and Hygiene* 58.3 (1998), pp. 277–282.
- [58] Florence Fouque et al. “*Aedes aegypti* survival and dengue transmission patterns in French Guiana.” In: *Journal of Vector Ecology* 31.2 (2006), pp. 390–399.
- [59] Lauren B Carrington et al. “Fluctuations at a low mean temperature accelerate dengue virus transmission by *Aedes aegypti*.” In: *PLoS Negl Trop Dis* 7.4 (2013), e2190.

Appendices

A. Equations	56
B. Algorithms & Code	60

A. Equations

A.1 Demography Model Equation Summary

Table 6: Equation list for demographic model

Table 6.A Host model transition probability equations.

Description	Equation
\mathbb{P} [Exposure]	$\lambda_H(t) = 1 - \exp\left(-\frac{\beta(t)I_G(t)}{N_H(t)}\right)$
\mathbb{P} [Infection]	$\sigma_H = 1 - \exp(-\hat{\sigma}_H)$
\mathbb{P} [Recovery]	$\gamma_H = 1 - \exp(-\hat{\gamma}_H)$
\mathbb{P} [Death]	$\mu_H = 1 - \exp(-\hat{\mu}_H)$
Birth recruitment	$B_H(t) \sim \text{Pois}\left(N_H(t)\hat{\phi}_H(t)\right)$
New Exposed, Dead	$[W_H(t), X_H^S(t)] \sim \text{Multinom}(S_H(t), (1 - \mu_H)\lambda_H, \mu_H)$
New Infected, Dead	$[V_H(t), X_H^E(t)] \sim \text{Multinom}(E_H(t), (1 - \mu_H)\sigma_H, \mu_H)$
New Recovered, Dead	$[U_H(t), X_H^I(t)] \sim \text{Multinom}(I_H(t), (1 - \mu_H)\gamma_H, \mu_H)$
Cases Reported	$c_H(t) C \sim \text{Normal}(\rho C, \rho(1 - \rho)C + (\psi \rho C)^2)$
Dead	$X_H^R(t) \sim \text{Binom}(R_H(t), \mu_H)$

Table 6.B Vector model transition probability equations

Description	Equation
\mathbb{P} [$J_G \rightarrow S_G$]	$\rho_G = 1 - \exp(-\tau)$
\mathbb{P} [$S_G \rightarrow E_G$]	$\lambda_G(t) = 1 - \exp\left(-\frac{\beta(t)(I_H(t) + \eta)\zeta(\epsilon)}{N_H(t)}\right)$
\mathbb{P} [$E_G \rightarrow I_G$]	$\sigma_G = 1 - \exp(-\hat{\sigma}_G)$
Juvenile death rate	$\mu_J(t) = 1 - \exp(-\hat{\mu}_J - \alpha J_G(t))$
Adult death rate	$\mu_G = 1 - \exp(-\check{\mu}_G)$
Egg recruitment	$B_G(t) \sim \text{Pois}\left(\hat{\phi}_G \hat{\theta}_G N_G(t)\right)$
New Susceptible, Dead	$[Z_G(t), X_G^J(t)] \sim \text{Multinom}(J_G(t), (1 - \mu_J(t))\rho_G, \mu_J(t))$
New Exposed, Dead	$[W_G(t), X_G^S(t)] \sim \text{Multinom}(S_G(t), (1 - \mu_G)\lambda_G(t), \mu_G)$
New Infected, Dead	$[U_G(t), X_G^E(t)] \sim \text{Multinom}(E_G(t), (1 - \mu_G)\sigma_G, \mu_G)$
Dead	$X_G^I(t) \sim \text{Binom}(I_G(t), \mu_G)$

A.2 Vector Lifespan

We calculate $(\hat{\mu}_J)^{-1}$ such that the probability of vector survival is 80%, assuming average adult vector lifespan, $(\check{\mu}_G)^{-1}$, is 10 days:

$$\exp(-10 * \hat{\mu}_J) = 0.8 \Rightarrow (\hat{\mu}_J)^{-1} = \left[-\frac{10}{\ln(0.8)} \right] = 45. \quad (29)$$

A.3 Density-Dependent Competition

To find the strength of juvenile competition factor α that maintains a user-defined total juvenile population, we want to satisfy $N_G(t+1) = N_G(t) \equiv N_G^*$ and $J_G(t+1) = J_G(t) \equiv J_G^*$. The change in total adult mosquito populations comes from the previous population combined with the incoming susceptibles and removal of adult deaths:

$$\begin{aligned} N_G(t+1) &= N_G(t) + \exp(-\hat{\mu}_J - \alpha J_G(t))(1 - \exp(-\tau))J_G(t) \\ &\quad - (1 - \exp(-\check{\mu}_G))N_G(t) \\ \Rightarrow \quad 0 &= \exp(-\hat{\mu}_J - \alpha J_G^*(1 - \exp(-\tau)))J_G^* - (1 - \exp(-\check{\mu}_G))N_G^*. \end{aligned} \quad (30)$$

Similarly, the change in juveniles results from the previous number of juveniles combined with incoming of new eggs and the removal of juvenile deaths:

$$\begin{aligned} J_G(t+1) &= J_G(t) + \hat{\psi}_G \hat{\theta}_G N_G(t) - (1 - \exp(-\hat{\mu}_J - \alpha J_G(t) - \tau))J_G(t) \\ \Rightarrow \quad 0 &= \hat{\psi}_G \hat{\theta}_G N_G^* - (1 - \exp(-\hat{\mu}_J - \alpha J_G^* - \tau))J_G^*. \end{aligned} \quad (31)$$

Since all other parameters are given, to find α it remains to solve two equations (30) and (31) for unknowns N_G^* and J_G^* . From (30),

$$N_G^* = \frac{\exp(-\hat{\mu}_J - \alpha J_G^*(1 - \exp(-\tau)))J_G^*}{(1 - \exp(-\check{\mu}_G))}. \quad (32)$$

Plugging (32) into (31) and solving for J_G^* gives

$$J_G^* = \frac{1}{\alpha} \left[\ln \left(\frac{\hat{\psi}_G \hat{\theta}_G (1 - \exp(-\tau)) + \exp(-\tau)}{(1 - \exp(-\check{\mu}_G))} \right) - \hat{\mu}_J \right]. \quad (33)$$

Then substituting (33) back into (32) and letting $\rho_G = 1 - \exp(-\tau)$ and $\mu_G = 1 - \exp(-\check{\mu}_G)$ for simplification:

$$\begin{aligned} N_G^* &= \frac{1 - \exp(-\tau) \ln \left(\frac{\hat{\psi}_G \hat{\theta}_G (1 - \exp(-\tau))}{(1 - \exp(-\check{\mu}_G))} + \exp(-\tau) - \hat{\mu}_J \right)}{\alpha (1 - \exp(-\check{\mu}_G)) \left(\frac{\hat{\psi}_G \hat{\theta}_G (1 - \exp(-\tau))}{(1 - \exp(-\check{\mu}_G))} + \exp(-\tau) \right)} \\ &= \frac{\rho_G \ln \left(\frac{\hat{\psi}_G \hat{\theta}_G \rho_G}{\mu_G} + (1 - \rho_G) - \hat{\mu}_J \right)}{\alpha \left(\hat{\psi}_G \hat{\theta}_G \rho_G + \mu_G (1 - \rho_G) \right)}. \end{aligned} \quad (34)$$

Finally solving the general form of α :

$$\alpha = \frac{\rho_G \ln \left(\frac{\hat{\psi}_G \hat{\theta}_G \rho_G}{\mu_G} + (1 - \rho_G) - \hat{\mu}_J \right)}{N_G \left(\hat{\psi}_G \hat{\theta}_G \rho_G + \mu_G (1 - \rho_G) \right)}. \quad (35)$$

Thus, to find the competition and initial juvenile population that keep the total mosquito population at equilibrium, one simply needs to plug the desired total juvenile population into (35) then input the resulting α into (33). Using an approximate 2:1 ratio of human to mosquitoes, we let $N_G^* = S_G(t_0) = 800,000$, which yields $\alpha = 4.2533576 \times 10^{-07}$ and $J_G^* = J_G(t_0) = 3,226,269$.

A.4 Measurement Model Sampling

The evaluation of a draw from the measurement density $f_{Y_n|X_n}(y_n|x_n;\theta)$ with normal cumulative distribution function $\Phi(\cdot, \mu, \sigma^2)$ is given by

$$\rho_H = \mathbb{P}[c | \rho, \psi, C] = \begin{cases} \Phi(c + \frac{1}{2}, \rho C_H, \rho(1 - \rho) C_H + (\psi \rho C_H)^2) \\ \quad - \Phi(c - \frac{1}{2}, \rho C_H, \rho(1 - \rho) C_H + (\psi \rho C_H)^2), & c > 0 \\ \Phi(c + \frac{1}{2}, \rho C_H, \rho(1 - \rho) C_H + (\psi \rho C_H)^2), & c = 0. \end{cases} \quad (36)$$

B. Algorithms & Code

B.1 IF2 Algorithm

model input: Simulators for $f_{X_0}(x_0; \theta)$ and $f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta)$; evaluator for $f_{Y_n|X_n}(y_n|x_n; \theta)$; data, $y_{1:N}^*$

algorithmic parameters: Number of iterations, M ; number of particles, J ; initial parameter swarm, $\{\Theta_j^0, j = 1, \dots, J\}$; perturbation density, $h_n(\theta|\varphi; \sigma)$; perturbation scale, $\sigma_{1:M}$

output: Final parameter swarm, $\{\Theta_j^M, j = 1, \dots, J\}$

1. For m in $1:M$
2. $\Theta_{0,j}^{F,m} \sim h_0(\theta|\Theta_j^{m-1}; \sigma_m)$ for j in $1:J$
3. $X_{0,j}^{F,m} \sim f_{X_0}(x_0; \Theta_{0,j}^{F,m})$ for j in $1:J$
4. For n in $1:N$
5. $\Theta_{n,j}^{P,m} \sim h_n(\theta|\Theta_{n-1,j}^{F,m}, \sigma_m)$ for j in $1:J$
6. $X_{n,j}^{P,m} \sim f_{X_n|X_{n-1}}(x_n|X_{n-1,j}^{F,m}; \Theta_j^{P,m})$ for j in $1:J$
7. $w_{n,j}^m = f_{Y_n|X_n}(y_n^*|X_{n,j}^{P,m}; \Theta_{n,j}^{P,m})$ for j in $1:J$
8. Draw $k_{1:J}$ with $P[k_j = i] = w_{n,i}^m / \sum_{u=1}^J w_{n,u}^m$
9. $\Theta_{n,j}^{F,m} = \Theta_{n,k_j}^{P,m}$ and $X_{n,j}^{F,m} = X_{n,k_j}^{P,m}$ for j in $1:J$
10. End For
11. Set $\Theta_j^m = \Theta_{N,j}^{F,m}$ for j in $1:J$
12. End For

Sequential Monte Carlo (SMC)
subroutine

B.2 pomp Model Configuration

Table 7: Mathematical notation for methods in `pomp` constructor.

pomp argument	Mathematical terminology
<code>rprocess</code>	Simulate from $f_{X_n X_{n-1}}(x_n x_{n-1};\theta)$
<code>dprocess</code>	Evaluate $f_{X_n X_{n-1}}(x_n x_{n-1};\theta)$
<code>rmeasure</code>	Simulate from $f_{Y_n X_n}(y_n x_n;\theta)$
<code>dmeasure</code>	Evaluate $f_{Y_n X_n}(y_n x_n;\theta)$
<code>initializer</code>	Simulate from $f_{X_0}(x_0;\theta)$
<code>t0</code>	t_0
<code>times</code>	$t_{1:N}$
<code>data</code>	$y_{1:N}^*$
<code>states</code>	$x_{0:N}$
<code>params</code>	θ

B.3 Source Code

Visit the repository https://bitbucket.org/nlevick/iqitos_thesis_pomp_code or email levick@unm.edu for example code.