STATISTICAL METHODS FOR FITTING DENGUE DISEASE MODELS, AND RELATED ISSUES

by

Yu-Ting Weng

M.S. in Applied Statistics, University of Pittsburgh, 2010

B.S. in Biological Science and Technology, National Chiao Tung

University, Taiwan, 2006

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Yu-Ting Weng

It was defended on

November 13rd

and approved by

Abdus S. Wahed, PhD, Associate Professor, Department of Biostatistics, Graduate

School of Public Health, University of Pittsburgh

Stewart J. Anderson, PhD, Professor, Department of Biostatistics, Graduate School of

Public Health, University of Pittsburgh

Satish Iyengar, PhD, Professor and Chair, Department of Statistics, Graduate School of

Public Health, University of Pittsburgh

Willem G. van Panhuis, MD, PhD, Assistant Professor, Department of Epidemiology,

Graduate School of Public Health, University of Pittsburgh

Shawn T. Brown, PhD, Director of Public Health Applications, Pittsburgh

Supercomputing Center, Carnegie Mellon University

Dissertation Director: **Abdus S. Wahed**, PhD, Associate Professor, Department of

Biostatistics, Graduate School of Public Health, University of Pittsburgh

STATISTICAL METHODS FOR FITTING DENGUE DISEASE MODELS, AND RELATED ISSUES

Yu-Ting Weng, PhD

University of Pittsburgh, 2014

Abstract:

Dengue is currently the fastest growing vector-borne disease which causes fever, headache, muscle aches, and other flu-like symptoms, affecting 50-100 million people worldwide yearly. Modeling dengue incidence over time is challenging because of multiple virus serotypes, high asymptomaticity, and the limited data availability. Different dengue modeling approaches have been explored in the public health literature such as economic models, agent-based (AB) models, and ordinary differential equation (ODE) models. ODE models are the standard to model dynamic systems involving interactions between various populations because of their solid mathematical/statistical foundation and ease of implementation in standard software packages. The assumptions of the homogeneity and perfect mixing of the ODE model, however, may not accurately represent the real world. On the other hand, AB models may lack the solid mathematical/statistical theory, but can model heterogeneity at the individual level. In the first part of this dissertation, we propose a simplified new ODE model (vSEIR) and compare this model with three existing ODE models. We also compare two discretization methods for initial value problems: derivative-free mesh adaptive direct search method with quadratic models (MADSQ) and derivative trust region (DTR) method. The simulation studies show that MADSQ can provide a better solution to the ODE compared to DTR when the parameter space has many local minima. We also demonstrate that the proposed vSEIR ODE model provides a better fit to the data than the other existing ODE models. In the second part of this dissertation, we validate a dengue **C**omputationa**L AR**thropod **A**gents

(CLARA) AB model, by comparing with its corresponding ODE model and the real world data. We not only show the similarity between the two models, but also contrast them. Our future plan is to continue to improve dengue ODE models by providing a stochastic version. Improved dengue models will provide public health researchers tools to better understand dengue disease outbreaks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First of all, I would like to express my acknowledgement to my advisors Dr. Shawn T. Brown and Dr. Abdus S. Wahed. Your help and guidance during the various stages of my dissertation is indispensable. Because you not only share your valuable and candid experiences and comments, but also have insisted on the high standards for the conducted research and quality of written contents, this dissertation thesis has been shaped into a better form. Thanks for your patience and kindness. I have appreciated our meetings which have cultivated me to think independently and critically.

I am also indebted to my committee members: Dr. Stewart J. Anderson, Dr. Satish Iyengar, and Dr. Willem G. van Panhuis. Because of your extremely useful and crucial suggestions, the depth of this dissertation thesis can be extensible.

I am grateful for Dr. Donald S. Burke, Dr. John J. Grefenstette, and Dr. Sally C. Morton who have provided the financial support via positions of graduate student research (GSR) and teaching assistant at the public health dynamic laboratory and the Department of Biostatistics throughout these years. I appreciate your feedback and discussions which have educated me to think maturely in different perspectives.

I also thank Dr. Nathan T.B. Stone (Pittsburgh Supercomputing Center) for discussing dengue simulation models, Dr. Scott Ritchie (James Cook University) and Dr. Gonzalo Vazquez-Prokopec (Emory University) for providing dengue fever instance data, Dr. David Swigon (University of Pittsburgh) for offering mathematical guidance, Dr. Jonathan Currie (Auckland University of Technology) for providing advice on OPTI software, Dr. Roberto Gomez (Pittsburgh Supercomputing Center) and Dr. Anirban Jana (Pittsburgh Supercomputing Center) for offering technical advice on the parallel simulation, and all computational support from Pittsburgh Supercomputing Center.

1.0   INTRODUCTION

Dengue fever is currently the world's fastest growing vector-borne disease. While more than 2.5 billion people living in areas of risk, around 50-100 million people are infected every year, mostly in urban and semi-urban areas [Halstead, 2007]. Although countries have implemented different vector control strategies [Yeap et al., 2011], these interventions have achieved only limited success, and the trend of the global spread is expanding [Horstick et al., 2010]. No licensed dengue vaccine is available for the public either, although several vaccine candidates are currently being evaluated in clinical studies [Guy et al., 2011]. Thus, global control of dengue fever is a major public health problem with significant economic, political, and social impact [Kyle and Harris, 2008] and developing different models to tackle the problem is the main task in this field. However, modeling the spread of dengue is challenging not only because it involves numerous complex factors such as the interactions between humans and mosquitoes, multiple coexisting virus serotypes, and high asymptomaticity of the initial infection [Kyle and Harris, 2008], but also because of the lack of existing data [Andraud et al., 2012].

Modeling dengue fever has been explored from different perspectives in the public health literature. From the economic point of view, the actual cost of dengue fever is very high as shown in a study from Puerto Rico [Halasa et al., 2012]. Using economic models, the impact of dengue fever was investigated [Beatty et al., 2011] and the study concluded that the economic literature is relatively sparse and results have often been conflicting due to the use of inconsistent assumptions. Additionally, the economic impact of dengue fever in Thailand was investigated by the disability-adjusted life years (DALYs)[Clark et al., 2005]. The authors suggested the governments and international funding agencies should consider giving equal priority to DF/DHF research, prevention, and control as diseases currently

1

receiving more resources. In addition, a Markov simulation model was developed to evaluate the potential health and economic value of administering a dengue vaccine to infants in Thailand [Lee et al., 2011].

If the dynamics of the disease outbreak simulation is the main interest, using the ordinary differential equation (ODE) or agent-based (AB) model approach, however, would be what we resort to. The standard approach is to use compartmental models involving ordinary differential equations for the human and the mosquito populations. Several ODE models with different sets of assumptions have been proposed [Andraud et al., 2012]. Using an ODE model with one virus serotype of dengue, for example, different biting rates of normal and infected female *Aedes aegypti* (*Ae. aegypti*) have been explored [Luz et al., 2011]. The results showed that increasing the biting rate can lead to increased numbers of primary and secondary infections in humans, therefore leading to more severe biennial epidemics. Various vector control strategies using the ODE model with deterministic time lags have been modeled [Atkinson et al., 2007] and the results showed that the proportional policy outperforms a release policy in which the released mosquito population is held constant. Additionally, in multiple-serotype models, symptomatic and asymptomatic compartments have been considered [Sciprom et al., 2007]. The authors assumed only virus transmission from symptomatic infected individuals to susceptible mosquitoes and concluded that the number of asymptomatic individuals could affect the two serotype equilibrium state. On the other hand, an increasingly popular approach is to applying agent-based models to observe the dynamics of the complex system. Chao et al. [2012] explored how dengue transmission dynamics are affected by the vaccination and suggested that children should be prioritized to receive vaccine. Another model which used high quality outbreak data and mosquito trapping data from Cairns was developed by Karl et al. [2014]. Their study suggested that the reason that the observed explosive outbreak of 2008/2009 was due to the a shorter virus strain-specific extrinsic incubation period, but not the warmer weather and increased human movement.

However, which approach is better? Which model can support a potential public health policy with more reliable explanations? Which model can be used to predict the future disease outbreak more accurately? Public health policy makers always face similar dilemma.

Table 1 summarizes the historical dengue outbreak data from 2001 to 2009 at Queensland, Australia. The existing strategies implement vector control interventions immediately if there is a new infected human case, despite the fact that we can see not all outbreaks have long duration. For example, the duration of the 2001 outbreak at Townsville was 3 weeks. This implies that if there is a model that can predict accurately the magnitude of the infected incidences, it would help the public health policy maker choose between the magnitude of the infected incidences and the intervention cost. The standard approach to modeling dengue incidence is to use ODE models. One limitation of the ODE approach is the assumption of perfect mixing between compartments and homogeneity. In contrast, we can imagine there are geographic barriers to humans and mosquitoes not being to be perfect mixing. Also, the ODE approach does not consider the spatial components. In contrast, AB models do not have these constraints. Hence, ODE and AB models need to be validated by reasonable methods.

Therefore, the objective of this dissertation is twofold. First, given the sparse data availability, we would like to develop a better model that can explain the dynamics of the disease outbreak. We propose a new ODE model (vSEIR) and compare this model with three existing ODE models. We also compare two discretization methods for initial value problems: derivative-free mesh adaptive direct search method with quadratic models (MADSQ) and derivative trust region (DTR) method. Then we would like to show which method can perform more efficiently when the parameter space has many local minima. Second, we would like to validate the CLARA AB model. We propose two methods to validate the model by comparing with its corresponding ODE model and the real world data. We hope these methods can not only be a robust method for validating the **C**omputational **AR**thropod **A**gents (CLARA) AB model, but also be easily applied to various parts of operations in the real world. Moreover, we hope our CLARA AB model is convincing and can be used to assist with policy making in the future. The completed work on Aim 1 and Aim 2 will be shown in Chapter 2 and 3, respectively. In the remaining part of this chapter, we describe the concepts behind the ODE and AB models.

1.1 ORDINARY DIFFERENTIAL EQUATIONS AND RELATED DENGUE MODELS

Using ordinary differential equation (ODE) models is a standard approach to describe complex systems in various fields. For example, consider modeling the human population growth over time where one is interested in the estimation of the human population at a certain time point. If we assume that the growth rate of the human population depends linearly on the human population itself, the model can be represented as:

$$\frac{dN^H(t)}{dt} = \theta_H N^H(t), \qquad N^H(0) = N_0$$

where $N^H(t)$ is the number of humans at time t, N_0 is the number of humans at time 0, and θ_H is a proportional parameter. Then the corresponding solution will be:

$$N^H(t) = N_0 e^{\theta_H t} \tag{1.1}$$

Because the human population can not indefinitely continue to grow exponentially, there are limits to the population growth. A more realistic assumption would be to let the growth rate of the human population become smaller when the human population becomes larger. Then the corresponding model can be derived as follows [Pearl, 1927, Tsoularis and Wallace, 2002]:

$$\frac{dN^H(t)}{dt} = \theta_H N^H(t)(1 - \frac{N^H(t)}{K}), \qquad N^H(0) = N_0$$

where $N^H(t)$ is the number of humans at time t, N_0 is the number of humans at time 0, θ_H is a proportional parameter, and K is the carrying capacity. Then the corresponding solution will be:

$$N^H(t) = \frac{KN_0}{(K - N_0)e^{-\theta_H t} + N_0} \tag{1.2}$$

In addition, ODE models can help us to understand the interactions between various populations. When dengue fever becomes the world's fastest growing vector-borne disease and has a huge disease impact, people are interested in the estimation of the disease outbreak at a certain future time point. When we understand how *Ae. aegypti* interact with humans

4

(more description in Chapter 2), a simplified SIR model without the exposed states has been developed as follows [Andraud et al., 2012]:

$$N^V(t) = S^V(t) + I^V(t)$$
$$\frac{dS^V(t)}{dt} = \eta_V - \frac{\beta_V}{N^H(t)} S^V(t) I^H(t) - \mu_V S^V(t)$$
$$\frac{dI^V(t)}{dt} = \frac{\beta_V}{N^H(t)} S^V(t) I^H(t) - \mu_V I^V(t)$$
$$N^H(t) = S^H(t) + I^H(t) + R^H(t)$$
$$\frac{dS^H(t)}{dt} = -\frac{\beta_H}{N^H(t)} S^H(t) I^V(t)$$
$$\frac{dI^H(t)}{dt} = \frac{\beta_H}{N^H(t)} S^H(t) I^V(t) - \delta_H I^H(t)$$
$$\frac{dR^H(t)}{dt} = \delta_H I^H(t)$$

where $N^V(t)$ is the total number of female *Ae. aegypti* at time t and $S^V(t)$ and $I^V(t)$ are the number of susceptible and infectious female *Ae. aegypti*, respectively at time t; Additionally, N^H is the total fixed number of humans at time t, which is calculated as a total of the number of susceptible, infectious, and recovered humans at time t denoted as $S^H(t)$, $I^H(t)$, and $R^H(t)$, respectively.

There are five parameters in the SIR model: η_V is the recruitment rate of female *Ae. aegypti* in the model with the fixed birth of the female *Ae. aegypti*, δ_H is the human recovery rate, μ_V is the female *Ae. aegypti* death rate, and β_V (β_H) is the virus transmission rate from humans to female *Ae. aegypti* (from female *Ae. aegypti* to humans).

On the other hand, when the aquatic stages of the *Ae. aegypti* are taken into consideration, a more complicated model has also been proposed as follows [Atkinson et al., 2007]:

$$\frac{dN^V(t)}{dt} = \gamma N^V(t - \tau_e)\frac{K - N^V(t - \tau_e)}{K} - \mu_V N^V(t)$$

$$\frac{dE^V(t)}{dt} = \frac{ac}{N^H}(N^V(t) - E^V(t) - I^V(t))I^H(t)$$
$$- e^{-\mu_V \tau_V}\frac{ac}{N^H}(N^V(t - \tau_V) - E^V(t - \tau_V) - I^V(t - \tau_V))I^H(t - \tau_V) - \mu_V E^V(t)$$

$$\frac{dI^V(t)}{dt} = e^{-\mu_V \tau_V}\frac{ac}{N^H}(N^V(t - \tau_V) - E^V(t - \tau_V) - I^V(t - \tau_V))I^H(t - \tau_V) - \mu_V I^V(t)$$

$$\frac{dE^H(t)}{dt} = \frac{ab}{N^H}(N^H - E^H(t) - I^H(t))I^V(t)$$
$$- e^{-\mu_H \tau_H}\frac{ab}{N^H}(N^H - E^H(t - \tau_H) - I^H(t - \tau_H))I^V(t - \tau_H) - \mu_H E^H(t)$$

$$\frac{dI^H(t)}{dt} = e^{-\mu_H \tau_H}\frac{ab}{N^H}(N^H - E^H(t - \tau_H) - I^H(t - \tau_H))I^V(t - \tau_H)$$
$$- \mu_H E^H(t) - (\gamma + \mu_H)I^H(t)$$

where $N^V(t)$ is the number of the total number of female *Ae. aegypti* at time t and $E^V(t)$ and $I^V(t)$ are the number of exposed and infectious female *Ae. aegypti*, respectively at time t; Additionally, N^H is the total fixed number of humans, which is calculated as a total of the number of susceptible, exposed, infectious, and recovered humans at time t denoted as $S^H(t)$, $E^H(t)$, $I^H(t)$, and $R^H(t)$, respectively.

There are 10 parameters in this more complicated model. γ is the birth rate of *Ae. aegypti* in the model with the varied birth of the female *Ae. aegypti*. a is the biting rate (number of bites per unit time), b is the probability that a bite from an infected *Ae. aegypti* will infect a susceptible human, and c is the probability that a susceptible *Ae. aegypti* is infected from biting an infected human. τ_e, τ_V, and τ_H are the eggs, female *Ae. aegypti*, and human incubation periods, respectively. μ_V and μ_H are the female *Ae. aegypti* and human death rates, respectively. K is a population parameter related to the carrying capacity of the larval population.

When we do not have data on aquatic stages of the *Ae. aegypti* and we recognize that longer incubation periods of dengue fever require the exposed states, we strike a happy medium between the simplified SIR model and the more complicated model mentioned above. The result of this compromise is the proposed vSEIR model described in Chapter 2.

1.2 AGENT-BASED MODELS AND THE COMPUTATIONAL ARTHROPOD AGENTS ABM

In contrast to the assumption of the homogeneous mixing between populations in ODE models, agent-based (AB) models try to model the heterogeneous interaction relationships at an individual level under different conditions. A typical agent-based model has three elements: agents, including their attributes and behaviors; the relationship between agents; how agents interact with the environment [Macal and North, 2010]. Then, the model will depend on the relationship between different types of agents and the interaction between agents and the environment at each time unit. Following these guidelines, the **C**omputational **AR**thropod **A**gents (CLARA) AB model has been developed.

CLARA is an AB model that represents both individual hosts (humans) and vectors (multiple life stages of *Ae. aegypti*: eggs, larvae, and the adults). The model infrastructure is shown in Figure 1. The human hosts, which were simulated from the 2000 census data in Australia, go to work or school (based on their age) for eight hours per day, return to their neighborhoods in the evening, and reside indoors at night. As a vector, their full life cycle is modeled by a variety of agents in the system. Eggs and larvae are stored in oviposition sites. Adult male mosquitoes seek to mate with females whenever possible until their death. Each adult female mosquito, as depicted in Figure 2, mating with adult male mosquitoes, moves based on either the human host volatiles when biting human hosts, or the volatiles emanating from oviposition sites when laying eggs. Given the spatial and temporal resolution of ABM, this model is well suited to explore heterogeneity in both intervention strategies and disease spread. The detailed model description is described elsewhere [Stone et al.].

However, how to validate an AB model is a challenge. We use two methods to validate the model by comparing with its corresponding ODE model and the real world data. The details are described in Chapter 3

1.3 FIGURES AND TABLES
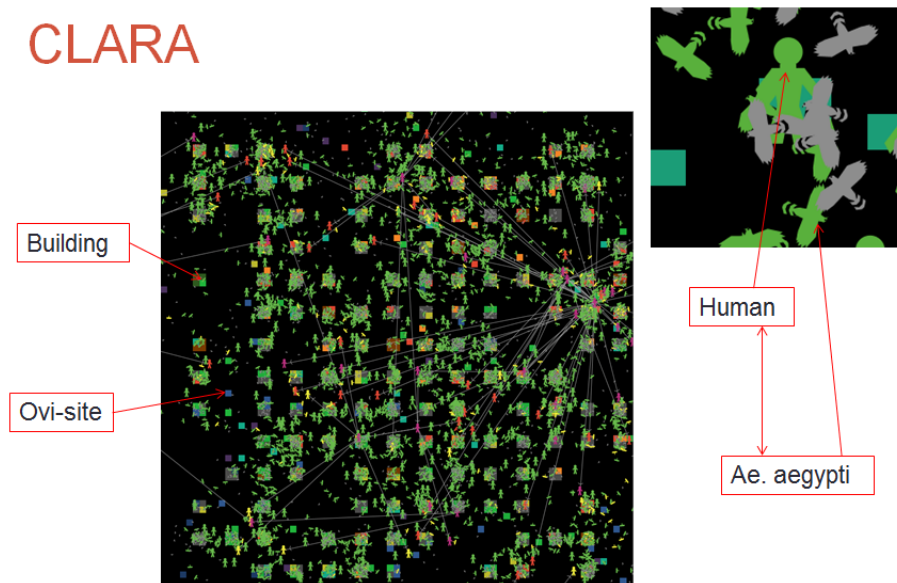
Figure 1: CLARA infrastructure
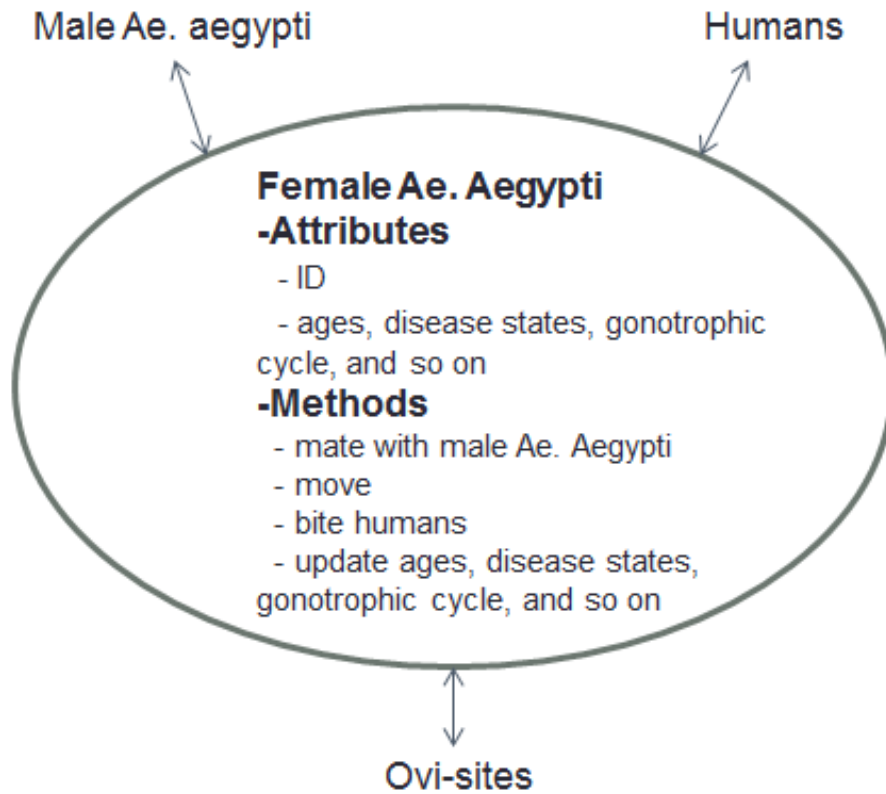
Figure 2: A female *Ae. aegypti* in CLARA

Table 1: Summary statistics for parts of Dengue outbreaks from 2001 to 2009 at Queensland, Australia

Year	Location	Reported Cases	Duration	Dengue Type
2001	Townsville	9	3 weeks	Dengue 2
2002	Cairns	2	3 weeks	Dengue 4
2003-4	Cairns	536	69 weeks	Dengue 2
2005	Townsville	18	22 weeks	Dengue 4
2006	Cairns	29	18 weeks	Dengue 2
2007	Townsville	46	13 weeks	Dengue 3
2008-9	Cairns, Townsville	931	29 weeks	Dengue 3

2.0 ESTIMATING DENGUE ORDINARY DIFFERENTIAL EQUATION MODELS WITH THE QUADRATIC MESH ADAPTIVE DIRECT SEARCH METHOD

2.1 INTRODUCTION

Ordinary differential equation (ODE) models are not only a standard approach to describe complex systems in many areas such as physics and biology, but also provide a way to understand the interactions between various populations [Fussmann et al., 2000]. In dengue studies, several ODE models with different sets of assumptions have been proposed [Andraud et al., 2012]. For example, using an ODE model with one serotype of dengue, a two-age-classes dengue transmission model with vaccination was explored [Luz et al., 2011]. The results showed that if there is an unintended vaccination of asymptomatic infectious children that effectively extends the infectious period, then the vaccination will cause a negative effect on disease prevention and treatment [Supriatna et al., 2008]. A severe Dengue Hemorrhagic Fever (DHF) compartment in a multi-serotype model was added [Nuraini et al., 2007]. In addition, a study, addressing n serotypes of dengue ($n = 2$-6), suggested that ADE may provide a competitive advantage to those serotypes that undergo enhancement compared with those that do not. This advantage increases with increasing numbers of cocirculating serotypes, but there are limits to the selective advantage provided by increasing levels of ADE [Cummings et al., 2005]. However, although various ODE models have been proposed and more complicated ODE models can be constructed, given the sparse data availability, we would like to know which model can really represent the complex real-world environment. Here, we consider four slightly different single serotype dengue models, compare their performance, and propose a better one given sparse data.

Estimation of ODE parameters from real-world data can be challenging because analytic solutions can be difficult to derive in most ODE models. Currently, there are two main approaches to fitting ODE models to real-world data. The first approach is to use the discretization methods for initial value problems, which means that the numerical solution of the ODE is constructed discretely at given time points [Bard, 1974]. The other approach to fitting the ODE model to real-world data, called the collocation methods, involves using basis function expansion. The general idea of this approach is to estimate the parameter by optimizing an objective function after the numerical solution of the ODE model in terms of a finite basis function expansion is obtained by satisfying the given equations at the given collocation points [Ramsay et al., 2007].

However, some issues exist in this spline-based method. Liang and Wu [2008] criticized that the penalized spline approaches require more efficient optimization techniques and complicated iterative computation algorithms to obtain an estimator. In addition, when the model includes a medium to large number of nonlinear parameters, or the data is incomplete or sparse, the results from the spline-based methods are not stable. In contrast, when using the discretization methods for initial value problems, the explicit information about the analytical or numerically approximated gradient or the Hessian matrix of objective function is likely to be unavailable or unreliable. Under these circumstances, the direct search method, one of the derivative free discretization methods for initial value problems, may be a reasonable alternative for finding a global minimizer of a real-valued objective function, even for problems in which the cost function is discontinuous, nondifferentiable, or stochastic [Conn et al., 2009].

The Nelder-Mead simplex algorithm, the most widely cited of the direct search methods, generates a new test position by extrapolating the behavior of the objective function measured at each test point, and replaces some of the parameters with new positions at the next iteration [Nelder and Mead, 1965]. However, for dimensions of parameters higher than 1, its convergence is uncertain [Lagarias et al., 1998]. For instance, it could converge to a nonminimizer even in a strictly convex unconstrained two-dimensional minimization problem [McKinnon, 1998].

Considering the shortcomings of these methods, we used a different method, namely, mesh adaptive direct search with quadratic models (MADSQ) to estimate the parameters in a proposed dengue ODE model. We used this method because it is derivative free, has much stronger convergence properties than the Nelder-Mead simplex algorithm, and performs better than the MADS algorithm without models [Conn and Le Digabel, 2013]. The MADSQ results are then compared with the results from a derivative trust region algorithm (DTR) included in the Intel Math Kernel Library (MKL) [Conn et al., 2000].

2.2   MODEL FORMULATION

Although the adult stage of *Ae. aegypti* mosquitoes cannot be represented separately from their stages of the aquatic life cycle, people attempted not to model the aquatic stages but only to develop a simplified continuous ODE model to reflect the interaction of humans with *Ae. aegypti* [Bailey, 1975]. An understanding of how *Ae. aegypti* interacts with humans is briefly described before developing a simplified one serotype dengue ODE model.

Female *Ae. aegypti* lay eggs that develop into larvae during an emergence period with a relatively high survival probability. Larvae feed on microorganisms and particulate organic matter, and change into pupae during an emergence period with a density-dependent survival probability. Pupae transform into adult *Ae. aegypti*. The entire aquatic life cycle lasts approximately 10-14 days until emerging into adult *Ae. aegypti*.

Female *Ae. aegypti* seek blood after they have mated with male *Ae. aegypti*. If the target humans are infected, the dengue virus transmits from the humans to the female *Ae. aegypti* given a certain probability, and the female *Ae. aegypti* become infectious in 10 to 14 days. On average, adult female *Ae. aegypti* can live about 10 days. Similarly, humans infected by the infectious female *Ae. aegypti* become infectious 4-7 days after exposure and remain infectious for up to 12 days. For up to 6 months, they are not susceptible to other dengue virus infections (i.e. serotypes have cross-immunity) and then become susceptible to the other serotypes of the dengue virus. Generally, humans can be infected with dengue fever twice during their lifespan although tertiary infections have been observed.

We make the following assumptions in our dengue ODE model

a) There are enough male *Ae. aegypti* for successful mating with female *Ae. aegypti*.

b) The birth of the female *Ae. aegypti* is a fixed number or a number which depends on the total number of the female *Ae. aegypti*.

c) Infected female *Ae. aegypti* go through susceptible (S), exposed (E), and infectious (I) states. No recovered state is assumed.

d) No extra vector control intervention parameters are added into the model due to the lack of statistical identifiability.

e) All infected humans are symptomatic. These infected humans go through susceptible (S), exposed (or not) (E), infectious (I), and recovered (R) states.

f) No human mortality exists due to the short time spans of observation.

g) The initial condition of the ODE model is known.

In addition, only one serotype of dengue virus is modeled to compare to DENV-2 or DENV-3 infected humans in our real-world data from 2003 and 2008, respectively.

The standard dengue model is derived from the SIR (Susceptible-Infectious-Recovered) infectious disease model [Andraud et al., 2012]. We form four different models by considering the fixed (f) or varied (v) birth of the female *Ae. aegypti*, and by including/excluding the exposed (E) state. These are denoted by fSIR [Andraud et al., 2012], vSIR [Derouich et al., 2003], fSEIR [Syafruddin and Noorani, 2012], and vSEIR, respectively. The general transition between various states can be represented using a vSEIR compartmental model shown in Figure 1. Basically, susceptible, exposed, infected female mosquitoes die at the same constant rate. Susceptible (infected) mosquitoes interact with infected (susceptible) humans, spreading the dengue virus. Natural transitions from S to E to I and to R in humans and from S to E to I in female *Ae. aegypti* occur at various rates. The vSEIR model depicted in Figure 3 can be presented as a system of ODEs as follows:

$$N^V(t) = S^V(t) + E^V(t) + I^V(t)$$

$$\frac{dS^V(t)}{dt} = \theta_V N^V(t) - \frac{\beta_V}{N^H(t)} S^V(t) I^H(t) - \mu_V S^V(t)$$

$$\frac{dE^V(t)}{dt} = \frac{\beta_V}{N^H(t)} S^V(t) I^H(t) - \gamma_V E^V(t) - \mu_V E^V(t)$$

$$\frac{dI^V(t)}{dt} = \gamma_V E^V(t) - \mu_V I^V(t)$$

$$N^H(t) = S^H(t) + E^H(t) + I^H(t) + R^H(t)$$

$$\frac{dS^H(t)}{dt} = -\frac{\beta_H}{N^H(t)} S^H(t) I^V(t)$$

$$\frac{dE^H(t)}{dt} = \frac{\beta_H}{N^H(t)} S^H(t) I^V(t) - \gamma_H E^H(t)$$

$$\frac{dI^H(t)}{dt} = \gamma_H E^H(t) - \delta_H I^H(t)$$

$$\frac{dR^H(t)}{dt} = \delta_H I^H(t)$$

where $N^V(t)$ is the number of the total number of female *Ae. aegypti* at time t and $S^V(t)$, $E^V(t)$, and $I^V(t)$ are the number of susceptible, exposed, and infectious female *Ae. aegypti*, respectively at time t; Additionally, N^H is the total fixed number of humans at time t, which is calculated as a total of the number of susceptible, exposed, infectious, and recovered humans at time t denoted as $S^H(t)$, $E^H(t)$, $I^H(t)$, and $R^H(t)$, respectively.

There are seven parameters in the SEIR model. θ_V is the birth rate of *Ae. aegypti* in the model with the varied birth of the female *Ae. aegypti* (or η_V is the recruitment rate of female *Ae. aegypti* in the model with the fixed birth of the female *Ae. aegypti*). γ_V (μ_V) is female *Ae. aegypti* infection (death) rate, γ_H (δ_H) is the human infection (recovery) rate, and β_V (β_H) is the virus transmission rate from humans to female *Ae. aegypti* (from female *Ae. aegypti* to humans). In contrast, there are only five parameters in the SIR model (no γ_V and γ_H).

Traditionally, while considering the vector control intervention, a model with a breakpoint (before and after the vector control intervention) is considered. That is, some parameters (θ_V or η_V, μ_V, β_V, and β_H) should have different values before and after the vector control intervention. However, the data before the vector control are too limited to estimate

parameters accurately. Thus, we take an alternative approach where we assume these parameters have a combined effect across the whole episode of dengue outbreak and directly fit the model without a breakpoint in the data.

2.3   MODEL PROPERTIES

2.3.1   Stability Analysis

The properties of equilibria and their stability are analyzed separately for three cases: $\theta_V = \mu_V$, $\theta_V < \mu_V$, and $\theta_V > \mu_V$.

2.3.1.1 $\theta_V = \mu_V$:   Because $\dfrac{dN^V(t)}{dt} = 0$ and $\dfrac{dN^H(t)}{dt} = 0$, N^V and N^H are constant. Thus, the vSEIR model can be reformulated as follows:

$$
\begin{aligned}
\frac{dE^V(t)}{dt} &= \frac{\beta_V}{N^H}(N^V - E^V(t) - I^V(t))I^H(t) - \gamma_V E^V(t) - \mu_V E^V(t) \\
\frac{dI^V(t)}{dt} &= \gamma_V E^V(t) - \mu_V I^V(t) \\
\frac{dE^H(t)}{dt} &= \frac{\beta_H}{N^H}(N^H - E^H(t) - I^H(t) - R^H(t))I^V(t) - \gamma_H E^H(t) \qquad (1) \\
\frac{dI^H(t)}{dt} &= \gamma_H E^H(t) - \delta_H I^H(t) \\
\frac{dR^H(t)}{dt} &= \delta_H I^H(t)
\end{aligned}
$$

The dynamical properties of the system (1) are described by the following results:

Theorem 1. *Let $D = \{(E^V, I^V, E^H, I^H, R^H) \epsilon \bar{\mathbb{R}}_+^5, 0 \le E^V + I^V \le N^V, 0 \le E^H + I^H + R^H \le N^H\}$. Assuming that the initial conditions lie in D, the system (1) has a unique solution that remains in D for $t \ge 0$.*

Proof. For system (1), we find that $E^V = 0$ implies $\dfrac{dE^V}{dt} \ge 0$; $I^V = 0$ implies $\dfrac{dI^V}{dt} \ge 0$; $E^H = 0$ implies $\dfrac{dE^H}{dt} \ge 0$; $I^H = 0$ implies $\dfrac{dI^H}{dt} \ge 0$; $R^H = 0$ implies $\dfrac{dR^H}{dt} \ge 0$. In addition, $E^V + I^V = N^V$ implies $\dfrac{dE^V}{dt} + \dfrac{dI^V}{dt} \le 0$, and $E^H + I^H + R^H = N^H$ implies $\dfrac{dE^H}{dt} + \dfrac{dI^H}{dt} + \dfrac{dR^H}{dt} \le 0$. Thus, the set D is invariant. $\qquad \square$

Theorem 2. *The system (1) has a set of equilibrium points, $x_{dfe} = (0,0,0,0,R^H)$ and $0 \leq R^H \leq N^H$.*

Proof. In order to find the solutions in the equilibrium state, we set all derivatives are equal to zero. Then, the solutions of the disease free equilibrium are $x_{\text{dfe}} = \{ E^V, I^V, E^H, I^H, R^H \} = \{ 0, 0, 0, 0, R^H \}$ and $0 \leq R^H \leq N^H$. $\square$

Theorem 3. *The subset of equilibrium points, $x_{dfe} = (0,0,0,0,R^H)$, where $\bar{R}^H < R^H \leq N^H$ and $\bar{R}^H$ is the critical point ($\bar{R}^H = N^H - \dfrac{(\gamma_V + \mu_V)\mu_V \delta_H (N^H)^2}{\gamma_V \beta_V \beta_H N^V}$) is comprised of equilibria that are locally stable in the sense of Lyapunov.*

Proof. See the Appendix A. $\square$

2.3.1.2 $\theta_V < \mu_V$: Because $\dfrac{dN^H(t)}{dt} = 0$, N^H is constant. Thus, the vSEIR model can be reformulated as follows:

$$
\begin{aligned}
\frac{dS^V(t)}{dt} &= \theta_V N^V(t) - \frac{\beta_V}{N^H} S^V(t) I^H(t) - \mu_V S^V(t) \\
\frac{dE^V(t)}{dt} &= \frac{\beta_V}{N^H} S^V(t) I^H(t) - \gamma_V E^V(t) - \mu_V E^V(t) \\
\frac{dI^V(t)}{dt} &= \gamma_V E^V(t) - \mu_V I^V(t) \\
\frac{dE^H(t)}{dt} &= \frac{\beta_H}{N^H}(N^H - E^H(t) - I^H(t) - R^H(t)) I^V(t) - \gamma_H E^H(t) \\
\frac{dI^H(t)}{dt} &= \gamma_H E^H(t) - \delta_H I^H(t) \\
\frac{dR^H(t)}{dt} &= \delta_H I^H(t)
\end{aligned}
\tag{2}
$$

The dynamical properties of the system (2) are described by the following results:

Theorem 4. *Let $D = \{(S^V, E^V, I^V, E^H, I^H, R^H) \in \bar{\mathbb{R}}_+^6, 0 \leq S^V + E^V + I^V \leq N^V, 0 \leq E^H + I^H + R^H \leq N^H\}$. Assuming that the initial conditions lie in D, the system (2) has a unique solution that remains in D for $t \geq 0$.*

Proof. For system (2), we find that $S^V = 0$ implies $\dfrac{dS^V}{dt} \geq 0$; $E^V = 0$ implies $\dfrac{dE^V}{dt} \geq 0$; $I^V = 0$ implies $\dfrac{dI^V}{dt} \geq 0$; $E^H = 0$ implies $\dfrac{dE^H}{dt} \geq 0$; $I^H = 0$ implies $\dfrac{dI^H}{dt} \geq 0$; $R^H = 0$

16

implies $\dfrac{dR^H}{dt} \geq 0$. In addition, $S^V + E^V + I^V = N^V$ implies $\dfrac{dS^V}{dt} + \dfrac{dE^V}{dt} + \dfrac{dI^V}{dt} \leq 0$, and $E^H + I^H + R^H = N^H$ implies $\dfrac{dE^H}{dt} + \dfrac{dI^H}{dt} + \dfrac{dR^H}{dt} \leq 0$. Thus, the set D is invariant. $\quad\square$

Theorem 5. *The system (2) has a set of equilibrium points, $x_{dfe} = (0,0,0,0,0,R^H)$ and $0 \leq R^H \leq N^H$.*

Proof. In order to find the solutions in the equilibrium state, we set all derivatives are equal to zero. Then, the solutions of the disease free equilibrium are $x_{\text{dfe}} = \{ S^V, E^V, I^V, E^H, I^H, R^H \} = \{ 0, 0, 0, 0, 0, R^H \}$ and $0 \leq R^H \leq N^H$. $\quad\square$

Theorem 6. *The set of equilibrium points, $x_{dfe} = (0,0,0,0,0,R^H)$, where $0 \leq R^H \leq N^H$ is comprised of equilibria that are locally stable in the sense of Lyapunov.*

Proof. The Jacobian of the system (2) at equilibrium points x_{dfe} is given by:

$$
J(x_{\text{dfe}}) = \begin{pmatrix}
-\gamma_V - \mu_V & 0 & 0 & 0 & 0 & 0 \\
\gamma_V & -\mu_V & 0 & 0 & 0 & 0 \\
0 & \dfrac{\beta_H(N^H - R^H)}{N^H} & -\gamma_H & 0 & 0 & 0 \\
0 & 0 & \gamma_H & -\delta_H & 0 & 0 \\
\theta_V & \theta_V & 0 & 0 & (\theta_V - \mu_V) & 0 \\
0 & 0 & 0 & \delta_H & 0 & 0
\end{pmatrix}
$$

Since the Jacobian is a triangular matrix, all eigenvalues are found on the diagonal of the Jacobian: 0, $-(\gamma_V + \mu_V)$, $-\mu_V$, $-\gamma_H$, $-\delta_H$, and $(\theta_V - \mu_V)$. Since $\theta_V < \mu_V$, all eigenvalues except 0 are real negative. The eigenvector corresponding to $\lambda = 0$, $\{ E^V, I^V, E^H, I^H, S^V, R^H \}^T = \{ 0, 0, 0, 0, 0, 1 \}^T$, is the vector tangential to the equilibrium manifold. Thus, each equilibrium point is locally asymptotically stable within the invariant set of the dynamical system containing that equilibrium point. In addition, it follows that all equilibria are Lyapunov stable. $\quad\square$

17

2.3.1.3 $\theta_V > \mu_V$: Because $\dfrac{dN^H(t)}{dt} = 0$, N^H is constant. Thus, the vSEIR model can be reformulated as follows:

$$\frac{dS^V(t)}{dt} = \theta_V N^V(t) - \frac{\beta_V}{N^H} S^V(t) I^H(t) - \mu_V S^V(t)$$

$$\frac{dE^V(t)}{dt} = \frac{\beta_V}{N^H} S^V(t) I^H(t) - \gamma_V E^V(t) - \mu_V E^V(t)$$

$$\frac{dI^V(t)}{dt} = \gamma_V E^V(t) - \mu_V I^V(t) \tag{3}$$

$$\frac{dE^H(t)}{dt} = \frac{\beta_H}{N^H}(N^H - E^H(t) - I^H(t) - R^H(t))I^V(t) - \gamma_H E^H(t)$$

$$\frac{dI^H(t)}{dt} = \gamma_H E^H(t) - \delta_H I^H(t)$$

$$\frac{dR^H(t)}{dt} = \delta_H I^H(t)$$

The dynamical properties of the system (3) are described by the following results:

Theorem 7. *Let* $D = \{(S^V, E^V, I^V, E^H, I^H, R^H) \epsilon \bar{\mathbb{R}}_+^6, 0 \le E^H + I^H + R^H \le N^H\}$. *Assuming that the initial conditions lie in D, the system (3) has a unique solution that remains in D for* $t \ge 0$.

Proof. For system (3), we find that $S^V = 0$ implies $\dfrac{dS^V}{dt} \ge 0$; $E^V = 0$ implies $\dfrac{dE^V}{dt} \ge 0$; $I^V = 0$ implies $\dfrac{dI^V}{dt} \ge 0$; $E^H = 0$ implies $\dfrac{dE^H}{dt} \ge 0$; $I^H = 0$ implies $\dfrac{dI^H}{dt} \ge 0$; $R^H = 0$ implies $\dfrac{dR^H}{dt} \ge 0$. In addition, $E^H + I^H + R^H = N^H$ implies $\dfrac{dE^H}{dt} + \dfrac{dI^H}{dt} + \dfrac{dR^H}{dt} \le 0$. Thus, the set D is invariant. $\qquad\square$

Theorem 8. *The system (3) has a set of equilibrium points,* $x_{dfe} = (0,0,0,0,0,R^H)$ *and* $0 \le R^H \le N^H$.

Proof. In order to find the solutions in the equilibrium state, we set all derivatives are equal to zero. Then, the solutions of the disease free equilibrium are $x_{\text{dfe}} = \{\ S^V,\ E^V,\ I^V,\ E^H,\ I^H,\ R^H\ \} = \{\ 0,\ 0,\ 0,\ 0,\ 0,\ R^H\ \}$ and $0 \le R^H \le N^H$. $\qquad\square$

Theorem 9. *The set of equilibrium points,* $x_{dfe} = (0,0,0,0,0,R^H)$, *where* $0 \le R^H \le N^H$, *is not locally stable in the sense of Lyapunov.*

Proof. The Jacobian of the system (3) at equilibrium points x_{dfe} is given by:

$$J(x_{\text{dfe}}) = \begin{pmatrix} -\gamma_V - \mu_V & 0 & 0 & 0 & 0 & 0 \\ \gamma_V & -\mu_V & 0 & 0 & 0 & 0 \\ 0 & \dfrac{\beta_H(N^H - R^H)}{N^H} & -\gamma_H & 0 & 0 & 0 \\ 0 & 0 & \gamma_H & -\delta_H & 0 & 0 \\ \theta_V & \theta_V & 0 & 0 & (\theta_V - \mu_V) & 0 \\ 0 & 0 & 0 & \delta_H & 0 & 0 \end{pmatrix}$$

Since the Jacobian is a triangular matrix, all eigenvalues are found on the diagonal of the Jacobian: 0, $-(\gamma_V + \mu_V)$, $-\mu_V$, $-\gamma_H$, $-\delta_H$, and $(\theta_V - \mu_V)$. Since $\theta_V > \mu_V$, there is a positive eigenvalue. Thus, none of the equilibrium points are locally stable. One can easily show that $N^V(t) \to \infty$ as $t \to \infty$. $\qquad\square$

2.3.2 Basic Reproduction Number, $\mathbb{R}_0$:

When $\theta_V = \mu_V$, the basic reproduction number $\mathbb{R}_0$ can be derived by the next generation operator approach Castillo-Chavez et al. [2002]. Using the notation of that paper, for system (1), we let X = (R^H), Y = (E^V, E^H), Z = (I^V, I^H), and $x_{\text{dfe}} = (X^*, 0, 0) \in \bar{\mathbb{R}}^5$. When assuming that $g(X^*, Y, Z) = 0$, Y = $\tilde{g}(X^*, Z) = (\tilde{g}_1(X^*, Z), \tilde{g}_2(X^*, Z))$ with

$$\tilde{g}_1(X^*, Z) = \frac{\beta_V}{N^H}(N^V - I^V)I^H\left(\frac{\beta_V}{N^H}I^H + \gamma_V + \mu_V\right)^{-1}$$

$$\tilde{g}_2(X^*, Z) = \frac{\beta_H}{N^H}(N^H - I^H - R^H)I^V\left(\frac{\beta_H}{N^H}I^V + \gamma_H\right)^{-1}$$

Then since $A = D_Z(X^*, \tilde{g}(X^*, Z), Z)|_{Z=0}$, we have

$$A = \begin{pmatrix} -\mu_V & \dfrac{\gamma_V\beta_V N^V}{N^H(\gamma_V + \mu_V)} \\ \dfrac{\beta_H}{N^H}(N^H - R^H) & -\delta_H \end{pmatrix}.$$

Additionally, $A = M - D$, with $M \geqslant 0$, and $D > 0$. Then,

$$M = \begin{pmatrix} 0 & \dfrac{\gamma_V\beta_V N^V}{N^H(\gamma_V + \mu_V)} \\ \dfrac{\beta_H}{N^H}(N^H - R^H) & 0 \end{pmatrix}, D = \begin{pmatrix} \mu_V & 0 \\ 0 & \delta_H \end{pmatrix}$$

19

giving

$$D^{-1} = \begin{pmatrix} \mu_V^{-1} & 0 \\ 0 & \delta_H^{-1} \end{pmatrix}.$$

Thus, in accord with Castillo-Chavez et al. [2002], the reproduction number for system (1) is $\mathbb{R}_0 = \rho(MD^{-1}) = \sqrt{\frac{\beta_V \beta_H \gamma_V (N^H - R^H) N^V(0)}{\mu_V (\gamma_V + \mu_V) \delta_H (N^H)^2}}$, where $\rho(C)$ denotes the spectral radius of a matrix C.

2.3.3   Expose State Excluded

In order to compare the vSEIR model to fSIR model, we reformulated the fSIR model and find the invariant set D and the set of the equilibrium points.

Since $\frac{dN^V(t)}{dt} = 0$ and $\frac{dN^H(t)}{dt} = 0$, it follows that N^V and N^H are constant. Thus, the fSIR model can be reformed as follows:

$$\begin{aligned} \frac{dI^V(t)}{dt} &= \frac{\beta_V}{N^H}(N^V - I^V(t))I^H(t) - \mu_V I^V(t) \\ \frac{dI^H(t)}{dt} &= \frac{\beta_H}{N^H}(N^H - I^H(t) - R^H(t))I^V(t) - \delta_H I^H(t) \qquad (4) \\ \frac{dR^H(t)}{dt} &= \delta_H I^H(t) \end{aligned}$$

Theorem 10. *Let $D = \{I^V, I^H, R^H) \, \epsilon \, \bar{\mathbb{R}}_+^3, \, 0 \le I^V \le N^V, \, 0 \le I^H + R^H \le N^H\}$. Assuming that the initial conditions lie in D, the system of equations for the fSIR model has a unique solution that remains in D for $t \ge 0$.*

Proof. For system (4), we find that $I^V = 0$ implies $\frac{dI^V}{dt} \ge 0$; $I^H = 0$ implies $\frac{dI^H}{dt} \ge 0$; $R^H = 0$ implies $\frac{dR^H}{dt} \ge 0$. In addition, $I^V = N^V$ implies $\frac{dI^V}{dt} \le 0$, and $I^H + R^H = N^H$ implies $\frac{dI^H}{dt} + \frac{dR^H}{dt} \le 0$. Thus, the set D is invariant. $\qquad \square$

Theorem 11. *The system (4) has a set of equilibrium points, $x_{dfe} = (0,0,R^H)$ and $0 \le R^H \le N^H$.*

Proof. In order to find the solutions in the equilibrium state, we set all derivatives are equal to zero. Then, the solutions of the disease free equilibrium are $x_{\text{dfe}} = \{ I^V, I^H, R^H \} = \{ 0, 0, R^H \}$ and $0 \le R^H \le N^H$. $\qquad \square$

Theorem 12. *The set of equilibrium points, $x_{dfe} = (0,0,R^H)$, where $\bar{R}^H < R^H \leq N^H$ and $\bar{R}^H$ is the critical point ($\bar{R}^H = N^H - \dfrac{(\mu_V \delta_H (N^H)^2}{\beta_V \beta_H N^V})$), is locally stable.*

Proof. The Jacobian of the system (4) at equilibrium points x_{dfe} is given by:

$$J(x_{\text{dfe}}) = \begin{pmatrix} -\mu_V & \dfrac{\beta_V N^V}{N^H} & 0 \\ \dfrac{\beta_H(N^H - R^H)}{N^H} & -\delta_H & 0 \\ 0 & \delta_H & 0 \end{pmatrix}$$

The characteristic polynomial is as follows:

$$-\lambda\{(-\mu_V - \lambda)(-\gamma_H - \lambda) - \frac{\beta_V \beta_H N^V (N^H - R^H)}{(N^H)^2}\}, 0 \leq R^H \leq N^H.$$

Therefore, the Jacobian has one zero eigenvalue and two non-zero eigenvalues. According to the criteria of Routh Hurwitz, the non-zero eigenvalues of $J(x_{\text{dfe}})$ have negative real parts if
$$\frac{\beta_V \beta_H N^V N^H}{(N^H)^2} - \mu_V \delta_H < \frac{\beta_V \beta_H N^V R^H}{(N^H)^2}.$$

Thus, the equilibrium point $(0,0,R^H)$ is stable if $\bar{R}^H = N^H - \dfrac{(\mu_V \delta_H (N^H)^2}{\beta_V \beta_H N^V} < R^H$. $\square$

We can take advantage of the Routh-Hurwitz criterion $a_0 > 0$ to compare the ranges of stability of the vSEIR and fSIR models. For the vSEIR model as derived above,

$$a_0 = \gamma_V \gamma_H \left[\frac{\gamma_V + \mu_V}{\gamma_V}\mu_V \delta_H - \frac{\beta_V \beta_H N^V (N^H - R^H)}{(N^H)^2}\right]$$

For the fSIR model,
$$a_0 = \mu_V \delta_H - \frac{\beta_V \beta_H N^V (N^H - R^H)}{(N^H)^2}$$

Because $\dfrac{\gamma_V + \mu_V}{\gamma_V} > 1$, the equilibrium points are stable for a larger range of R^H in vSEIR than fSIR. Thus, the presence of exposed state improves the stability of equilibria in these epidemiological models.

2.4 MODEL FITTING

Presence of parameters in various scales makes numerical implementation challenging, and hence we let the scale of all parameters are in $[0, 1]$ (η_V is rescaled here). Then we kept the range of those combined effect parameters in $[0, 1]$ since no supported studies can help us narrow the space of these parameters, but narrowed the space of other parameters based on some previous clinical studies [Gubler, 1998, Watts et al., 1987]. Finally, we assume that the virus transmission rate from humans to female *Ae. aegypti* is the same as the virus transmission rate from female *Ae. aegypti* to humans; the female *Ae. aegypti* and the human infection rates are known. Even so, fitting the above complex models to the data is challenging due to the sparse nature of the data. The data contains only the number of the new infectious humans (I_N^H) at different time points.

Our postulated model can be written as:

$$I_N^H(t) = m(\theta, t) + Z_t, \qquad t = 1,\dots,n$$

where $m(\theta, t) = \mathbb{E} I_N^H(t)$ is taken as the solution to the differential equation in terms of $I_N^H(t)$, and Z_t is the error term. For SEIR and SIR models, $m(\theta, t) = \gamma_H E^H(t)$ and $\frac{\beta_H}{N^H} S^H(t) I^V(t)$, respectively. The estimation of the parameter vector θ is carried out by minimizing the sum of squared error:

$$min_\theta \qquad f_\Theta(\theta) = \sum_{t=1}^{n} [I_N^H(t) - m(\theta, t)]^2$$

Note that $m(\theta, t)$ does not have a closed-form expression, and the solution $m(\theta, t)$ is highly non-linear in θ. There are two main approaches to fitting such models.

In the discretization methods, the parameters are estimated iteratively starting with an initial value, many of the methods developed under this approach are gradient-based methods. For example, the nonlinear least squares (NLS) methods, which use a numerical gradient and Hessian matrix of a real-valued objective function to search the optimal ODE parameter values, was proposed by Bard [1974]. Later, the asymptotic properties of the proposed estimators were proved by Xue et al. [2010]. Others have proposed Markov Chain Monte Carlo (MCMC) methods [Gelman et al., 2004], which assume observations follow some

density functions conditional on the parameters of interest and estimate the parameters from the Metropolis-Hastings or other sampling algorithms.

A spline-based method which uses a three-level criteria scheme was proposed by Ramsay et al. [2007]. While fixing the outer level criterion, the penalty parameter λ, the B spline coefficients are selected first at the inner level criterion. Then the optimal model parameter values are estimated at the middle level criterion. Not only its computational efficiency is as fast or faster than NLS and much faster than the MCMC method, but the bias and sampling variance estimation performance is also at least as good as other approaches such as MCMC and NLS. These concepts were further extended to partial differential equation estimation [Xun et al., 2013]. Simulation studies showed that the performance of their method is comparable to the MCMC and outperforms other available methods such as the two-stage method proposed by Marx and Eilers [2005].

Alternative approaches such as the Nelder-Mead approach are based on direct search methods. A generalized pattern search method (GPS) for unconstrained optimization was developed [Torczon, 1997]. This algorithm systematically evaluates test points that lie on a mesh or lattice centered at the current iteration, but local exploration of the space of variables is only restricted to a finite number of directions. The MADS algorithm, which was derived from GPS, was used to weaken the finite search direction restriction in GPS with much stronger convergence properties [Audet and Dennis JR., 2006] and was shown to perform better than GPS through simulations [Abramson and Audet, 2006]. The MADS algorithm is also more likely to achieve the optimal solution, and more efficient than the genetic algorithm (GA) [Das et al., 2011]. Later, the MADS algorithm was modified to form quadratic models at each iteration and the MADS algorithm with quadratic models (MADSQ) has been demonstrated to perform better than the MADS algorithm in many cases [Conn and Le Digabel, 2013].

2.4.1   Mesh Adaptive Direct Search with Quadratic Models

MADSQ is an iterative algorithm where points are generated (SEARCH and POLL steps) at each iteration based on the criteria presented below and the associated objective function

values $f_\Theta(\theta)$ are compared with the current best feasible objective function value $f_\Theta(\theta_k)$ found so far. The iteration will be ended when one of the termination criteria is satisfied. The algorithm in our case is as follows [Conn and Le Digabel, 2013]:

a) The initial point $\theta_0 \in \Theta^n$, the mesh size Δ_k^m, the poll size Δ_k^p, the radius factor ρ, and the quadratic basis $\{ 1, x_1, x_2, ..., x_n, \frac{x_1^2}{2}, \frac{x_2^2}{2}, ..., \frac{x_n^2}{2}, x_1x_2, ..., x_{n-1}x_n \}$ are given. Let the iteration k be 0 as well.

b) SEARCH step: construct an interpolation set Y first and then build a quadratic model $m_f(\theta)$ in $B_\infty(\theta, \rho\Delta_k^p)$. The trial point θ_k is obtained from the minimization of that quadratic model. Test θ_k to see if it is an improved mesh point.

c) POLL step: if the SEARCH step cannot find an improved mesh point, another quadratic model $m_f(\theta)$ in $B_\infty(\theta, \rho r)$, where r is the smallest radius including all Y. The trial point θ_k is obtained from the minimization of that quadratic model. Test θ_k if an improved mesh point.

d) Update Δ_{k+1}^m and Δ_{k+1}^p. Increment $k \leftarrow k + 1$ and go to the SEARCH and POLL step.

A detailed definition of the MADS frame work can be found in Audet and Dennis JR. [2006].

The objective function to be considered here is the least squares error $f_\Theta(\theta)$. The algorithm is implemented in C++ NOMAD which links to the *OPTI optimization toolbox* in MATLAB [Le Digabel, 2011]. Thus, we can implement in MATLAB and use this method via the *OPTI optimization toolbox*. The termination criteria are if either the maximal number of function evaluations is reached (10^5), the number of iterations the algorithm performs is reached (10^4), the step length is less or equal than 10^{-12}, the minimum distance between the current points at two consecutive iterations is less or equal than 10^{-6}, or the tolerance for the objective function is less or equal than 10^{-6}. The procedure provides the least square estimators for θ, $\hat{\theta}$. Since MADSQ is a direct search method, it does not provide a Hessian matrix. Therefore, for estimating the variance of $\hat{\theta}$, we resort to the sieve bootstrap developed in the time series literature to estimate variances for non-stationary time series models.

2.4.2 Sieve Bootstrap

The sieve bootstrap for nonstationary time series has been applied, since the dengue disease outbreak has a deterministic trend. Before the vector control intervention, numbers of new infectious cases increased; during the vector control intervention, numbers of new infectious cases decreased.

We consider the model, $Y_t = m(t) + Z_{t,n}$, where $\{m(t)\}_{t\in Z}$ is the deterministic differential equation function. Then, the sieve bootstrap procedure is the approximation of a stationary process with mean zero $Z_{t,n}$ by an autoregressive sequence of infinity lag with n time points.

The procedure can be implemented in the following steps:

a) Construct a kernel smoothers $\tilde{m}(t)$ for m(t), $t = [\delta n] + 1, \ldots, [(1 - \delta)n]$, $0 < \delta < 1/2$, since kernel smoothers are known to have larger bias near the end points. Then form the residuals, $\hat{Z}_{t,n} = Y_t - \tilde{m}(t)$.

b) The kernel smoothing estimator $\hat{m}$ has the bandwidth $\tilde{h} = ch^{5/9}$, where c is an arbitrary constant, and optimal bandwidth h is obtained from minimizing the following equation:

$$\sum_{t=1}^{n}[I_N^H(t) - \hat{m}(\theta, t)]^2.$$

c) Base on the Akaike information criterion (AIC) in a range of $[0, 10\log_{10}(n)]$ to decide on the order of p of the autoregressive model.

d) By using the Yule-Walker method [Yule, 1927, Walker, 1931], compute the coefficient of AR(p) model $\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n}$ based on $\hat{Z}_{t,n}$.

e) Form and center the residuals $\tilde{\varepsilon}_{t,n} = \hat{\varepsilon}_{t,n} - \sum_t \hat{\varepsilon}_{t,n}$, and $\hat{\varepsilon}_{t,n} = \sum_{j=0}^{p} \hat{\phi}_{t,n}\hat{Z}_{t-j,n}$, $t = p+[\delta n]+1, \ldots, [(1 - \delta)n]$.

f) Resample ε_t^* i.i.d. $\sim F_{\tilde{\varepsilon},n}$, the empirical cumulative distribution function of $\{\tilde{\varepsilon}_{t,n}\}_t$, $t = [\delta n] + 1, \ldots, [(1 - \delta)n]$.

g) Compute the residuals $\{Z_t^*\}_t$ based on $\sum_{j=0}^{p} \hat{\phi}_{t,n}Z_{t-j,n}^* = \varepsilon_{t,n}^*$, $Z_1^* = \varepsilon_1^*$. Then generate the bootstrap observations $Y_t^* = \hat{m}(t) + Z_t^*$, $t = [\delta n] + 1, \ldots, [(1 - \delta)n]$.

Here δ is the reciprocal of the time point plus one, and the detailed assumptions and definitions of the sieve bootstrap can be found in Bühlmann [1998].

2.5 SIMULATIONS

We carried out a simulation study to evaluate the performance of the MADSQ and the DTR algorithms in how they obtain the optimal solution in the vSEIR model. Data was generated from the model:

$$I_N^H(t) = m(\theta, t) + Z_t, \qquad t = 1, \ldots, n$$

with values of the true parameters $\theta = (\ \theta_V, \ \beta_V \ (= \beta_H), \ \mu_V, \ \delta_H \ )$ chosen based on the two data sets that motivated this study. We assumed γ_V and γ_H are known. We assumed the following initial conditions, $\{ \ S^V(0), E^V(0), I^V(0), S^H(0), E^H(0), I^H(0), R^H(0) \ \} = \{$ 22422, 0, 0, 2827, 1, 0, 0 $\}$ for scenarios I and II, and $\{$ 148019, 0, 0, 18669, 1, 0, 0 $\}$ for scenarios III and IV. The initial parameter values are calculated by taking the true values plus or minus ten percent of the true values.

For simplicity, instead of an autoregressive sequence of infinity lag, we generated Z_t from ARMA(1,1) model [Bühlmann, 1998]:

$$Z_t = 0.8 Z_{t-1} - 0.5 \varepsilon_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, n$$

where ε_t i.i.d. $\sim N(0,1)/\sqrt{1.2}$ (hereafter called *normal* for short), and $\text{Var}(Z_t)$ was rescaled to 0.16 (hereafter called *variance* for short). All simulation results are based on 200 realizations of such processes. For scenarios I and II, number of time points (n) was set to 124 (the same number as in the Cairns data set from 2003). For scenarios III and IV, number of time points (n) was set to 209 (the same number as in the Cairns data set from 2008). The estimates from different algorithms will be compared by the mean square error (MSE) of prediction defined by

$$MSE = \frac{\sum_{t=1}^n [m(\theta, t) - m(\bar{\theta}, t)]^2}{n}$$

where $\bar{\theta}$ is the mean of the estimators from 200 simulation realizations.

Table 2 displays the estimators and the Monte Carlo standard error (MCSE) of the parameters from 200 simulation realizations for different true parameter values. In scenarios I, the optimal mean estimator by MADSQ and DTR is unbiased and slightly biased, respectively. In scenarios II, optimal mean estimators by both algorithms are slightly biased.

However, the MADSQ algorithm estimates produce a prediction of new infectious humans closer to the number based on the true parameter value. The MSEs in these scenarios are smaller for MADSQ (MADSQ v.s. DTR: 0.001 v.s 0.089 in scenario I; 0.051 v.s. 0.158 in scenario II). The MCSE of the parameters is smaller for MADSQ than for DTR. In contrast, in scenarios III and IV, the optimal mean estimator by DTR is nearly unbiased. In addition, the data shows that the DTR algorithm estimates produce a prediction of new infectious humans closer to the number based on the true parameter value. The MSEs in these scenarios are smaller for DTR (DTR v.s. MADSQ: 0.613 v.s 3.474 in scenario III; 1.042 v.s. 6.723 in scenario IV). Furthermore, the MCSE of the parameters is smaller for DTR than for MADSQ in scenario III.

Table 3 displays the estimators and the MCSE of the parameters for 200 simulated data sets for two sets of initial parameter values (Init I and II) in scenario II. The results show that both algorithms depend on starting values to converge to the optimal estimator. Although optimal mean estimators by both algorithms are slightly biased, the MSEs in these two scenarios are smaller in MADSQ (MADSQ v.s. DTR: 0.051 v.s. 0.158 in Init I; 0.0001 v.s. 0.221 in Init II). In addition, the MCSE of the estimators is smaller in MADSQ.

The parameter space of the ODE system may not always have many local minima. For example, there are many local minima in scenario II but not in scenario III. We examined the parameter space in scenarios II (Figure 5.a) and III (Figure 5.b) in 2-D scenarios . When we varied two parameters (μ_V in [0,1] and δ_H) in [0.08,0.33] (μ_V: the death rate of the female *Ae. aegypti* and δ_H: the recovery rate of humans) but fixed others (θ_V and β_V), we could see multiple local minima ((θ_V, β_V) = (0.24, 0.1125), (0.23, 0.19), (0.22, 0.295)) in scenario II (Figure 5.a) but not in scenario III (Figure 5.b). This is one reason why the estimators by both algorithms were slightly biased in scenarios II but the estimators by DTR were nearly unbiased in scenario III.

2.6 ANALYSIS OF CAIRNS DATA

We fitted all four models (fSIR, fSEIR, vSIR, and vSEIR) to two data sets from two of the largest outbreaks that had occurred in the city of Cairns, Australia and the Cairns region of Australia (Cairns North, Clifton Beach, Whitfield, Parramatta Park, and Edge Hill) in 2003 and 2008, respectively. We describe the data sets first and discuss the modeling results in the following subsections.

2.6.1 The outbreak in 2003

In the 2003 outbreak (Figure 4.a), the onset of symptoms in the imported case occurred on January 22 in Parramatta Park (PP), but the dengue spread was not identified until March 2, and mosquito control measures were initiated the next day. A total of 383 laboratory-confirmed mild DENV-2 symptomatic cases were registered within urban Cairns over the 25-week epidemic period, but only the 233 cases in the neighborhood of PP over the 18-week epidemic period were included in this analysis. Differential equation models assume homogeneity and perfect mixing within compartments. If the confirmed symptomatic cases beyond PP regions were included, these assumptions would be more likely to be violated because there are some natural geographical barriers within urban Cairns that interfere with the interactions between different populations. The detailed data description can be found in Vazquez-Prokopec et al. [2010].

In 2003, the human population in PP was estimated to be 2,828 according to the data from the Australian Bureau of Statistics. Based on simulated models, we assumed the population of female *Ae. aegypti* was approximately 22,422 in PP during the outbreak period [Williams et al., 2013].

A clinical study reported that the extrinsic incubation period of dengue virus (EIP = $\frac{1}{\gamma_V}$) and the intrinsic incubation period in humans (IIP = $\frac{1}{\gamma_H}$) ranged between 10 and 14 days, and 3 and 14 days (average 4.5 to 7 days), respectively [Ritchie et al., 2013]. They also reported a total incubation period (TIP = EIP + IIP) of 17 days. Since we assumed the female *Ae. aegypti* (γ_V) and the human infection rates (γ_H) are known, we considered

two values of (EIP,IIP), satisfying TIP = 17 days, namely, (13,4) and (4,13), which cover the clinical range in the studies. These two assumptions will be denoted by A1 and A2. The model fits will be assessed based on the MSE defined by

$$MSE = \frac{\sum_{t=1}^{n}[I_N^H(t) - m(\hat{\theta}, t)]^2}{n}$$

where $\hat{\theta}$ is the estimator of the parameters.

For model comparisons, we let the dimension of the parameter space be the same $(\dim(\Theta) = 4)$ in all models and the results are shown in Table 4.a and Figure 6. We look at the fSIR model first. The top panel (a) of Table 4 provides parameter estimates for the model fits of the 2003 Cairns outbreak. The fSIR model could not fit the data well and the estimate of the recruitment rate of female *Ae. aegypti* η_V was 0. For the fSEIR model fit under the assumption A1, the virus transmission rate from female *Ae. aegypti* to humans (β_V) and the human recovery rate (δ_H) was slightly overestimated by the MADSQ algorithm compared to the DTR algorithm, whereas for the recruitment rate of female *Ae. aegypti* (η_V), the result was the opposite. The parameter estimates for the vSIR model were consistently higher for DTR compared to MADSQ, whereas for the vSEIR model, MADSQ provided larger estimates. In Figure 6, in the left panel, we show various model fits with the actual data. The MADSQ algorithm estimates in assumptions A1 and A2 produce similar new infectious humans, so the results were not added in. As we mentioned before, the fSIR model did not fit well to this data. We think it is because the total incubation period of the 2003 outbreak is longer (around 17 days). The longer incubation periods suggest the existence of the exposed state variables, which is not considered in the fSIR model and hence the fSIR model cannot fit the data well. However, for the vSIR and the vSEIR models, the fitted curve was closer to the observed values than the fSEIR model. The MSEs of the fSIR, fSEIR, vSIR, and vSEIR models fitted by the MADSQ method in the 2003 data set are 2.93, 2.58, 2.35, and 2.27, respectively, showing that vSIR and vSEIR models perform similarly for this data set.

We presented a 95% pointwise confidence interval for the number of new infections predicted using the vSEIR model under assumption A1 in Figure 5 where we let the bandwidth constant c to be 2.5. For comparison, we also constructed the confidence interval from the parametric bootstrap with the Poisson distribution [Chowell et al., 2007], and this 95%

pointwise confidence interval Figure (7.a) is wider than the one from the sieve bootstrap Figure (7.b).

2.6.2 The outbreak in 2008

In the 2008 outbreak (Figure 4.b), the onset of symptoms in the imported case occurred on November 5 in the Cairns North suburb; however, the dengue spread was not identified until November 28, and mosquito control measures were initiated on December 1. A total of 852 laboratory-confirmed mild DENV-3 symptomatic cases were registered within the Cairns region over the 30-week epidemic period, and all of the cases were included in this analysis because the geolocation data is confidential. The detailed data description can be found in Ritchie et al. [2013]. Compared to the outbreak in 2003, the epidemic covered a larger region. For this reason, the assumptions of the homogeneity and perfect mixing within compartments are more likely to be violated in the 2008 outbreak. For the initial settings, the human population in the Cairns region was estimated to be 18,669 in 2008, and we assumed the population of female *Ae. aegypti* was approximately 148,019 in the Cairns region during the outbreak period [Williams et al., 2013].

A clinical study reported that the total incubation period (TIP = EIP + IIP) is only 9 to 11 days [Ritchie et al., 2013]. After assuming TIP is 10 days, we considered two values of (EIP,IIP): (7,3) and (6,4), which cover the clinical range in the studies. These two assumptions will be denoted by A1 and A2.

For model comparisons, we let the dimension of the parameter space be the same ($\dim(\Theta)$ = 4) in all models and the results are shown in Table 4.b and Figure 6. We look at the fSIR model first. The bottom panel (b) of Table 4 provides parameter estimates for the model fits of the 2003 Cairns dataset. The fSIR model did not fit the data well. For the fSEIR model fit under the assumption A1, in half of cases, the virus transmission rate from female *Ae. aegypti* to humans (β_V) and the human recovery rate (δ_H) was slightly overestimated by the MADSQ algorithm compared to the DTR algorithm, whereas for the recruitment rate of female *Ae. aegypti* (η_V), the result was the opposite. In most instances, the parameter estimates for the vSIR model were slightly higher for DTR compared to MADSQ except

for the human recovery rate (δ_H), whereas for the vSEIR model, in every instance, the parameter estimates were reliably higher for DTR compared to MADSQ. In 6, in the right panel, we show various model fits with the actual data. The MADSQ algorithm estimates in assumptions A1 and A2 produce similar new infectious humans, so the results were not added in. As we mentioned before, the fSIR model did not fit well to this data because of the same reasons in the above subsection. However, for the vSEIR model, the fitted curve was closer to the observed values than others. The MSEs of the fSIR, fSEIR, vSIR, and vSEIR models fitted by the MADSQ method in the 2008 data set are 8.07, 6.29, 7.21, and 5.81, respectively, showing that vSEIR model perform the best for this data set.

2.7   DISCUSSION AND CONCLUSIONS

Scientists are used to applying ODE to model dynamic systems involving interactions between various populations. Here, we propose a new set of ODEs to model the incidence of dengue virus infection that elucidates the interaction between humans and mosquitoes throughout the life cycle of *Ae. aegypti*. When compared to other existing models (fSIR, fSEIR, and vSIR), both factors in the proposed model are necessary: the existence of the exposed state variables and the varied birth rate of the female *Ae. aegypti* which depends on the total number of the female *Ae. aegypti*. First, the incubation periods of female *Ae. aegypti* and humans are modeled by the exposed state variables in the model and longer incubation periods need the existence of the exposed state variables. Second, the implemented vector control interventions decrease the birth of the female *Ae. aegypti*. In other words, if the fixed birth rate models were used, the number of the total female *Ae. aegypti* would finally be at the equilibrium state, but in reality the vector control intervention would break the equilibrium state. Moreover, the MSE of the proposed model is the smallest among all models. Thus, after considering the results in both data sets qualitatively and quantitatively, we think the vSEIR model is a better choice. This model is in the class for which explicit information about the analytical or numerically approximated gradient or the Hessian matrix of objective function is unavailable or unreliable. We introduce mesh adaptive direct search

with quadratic models (MADSQ) to obtain the least squares estimator which avoids calculating a numerically approximated gradient or a Hessian matrix. To obtain the standard errors of estimates and confidence intervals, we think the sieve bootstrap for non-stationary time series is more reasonable than the parametric bootstrap because the correlation between the new infected humans at different time points is considered.

We have shown through simulation studies that the MADSQ algorithm performs more efficiently than the DTR algorithm when the parameter space has many local minima. However, the parameter space may not always have many local minima, so it is better to visualize the parameter space before deciding which optimization algorithm to use. Like many other non-linear optimization algorithms, the algorithm depends on a good starting value to converge to the optimal value, so it is better to choose more than one set of starting values and choose the set of the optimal value with the smallest mean squared error.

The MADSQ algorithm is available at *OPTI toolbox* linking to the NOMAD software in C++. It is easy to use and modify if necessary. However, the disadvantage of this approach is that it is computationally more expensive than the DTR algorithm, since it constructs quadratic models by interpolation, possibly leading to iterations where little or no improvement is seen. Parallel computing can significantly reduce the computational time.

The proposed method has the flexibility to be modified within intermediate steps or to be combined with other methods to improve efficiency. The variable neighborhood search (VNS) metaheuristic is incorporated into the search step of the MADS algorithm for exploring further test points to avoid trapping at local minima [Audet et al., 2008]. Compared to the classic MADS, this approach shows better simulation results, but the computational time is even longer. The optimization framework can be adapted to a large number of problems beyond ordinary differential equations. The MADS algorithm is extended to the stochastic constrained optimization and showed a better computational efficiency than Monte-Carlo schemes for problems with strict probabilistic constraints [Sankaran et al., 2010].
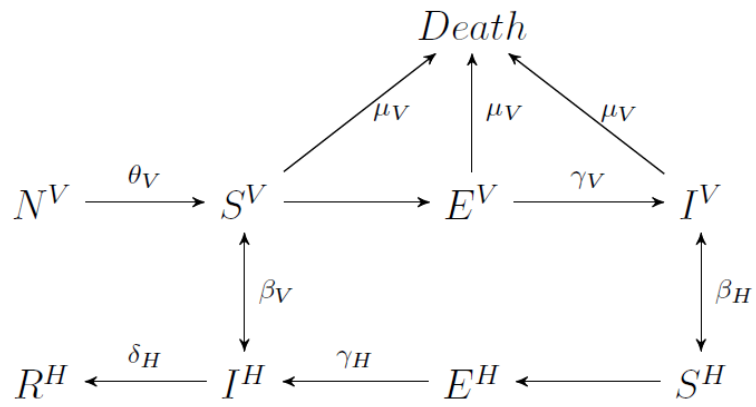
Figure 3: The diagram of the simplifed vSEIR model of the interaction between humans and mosquito in different stages
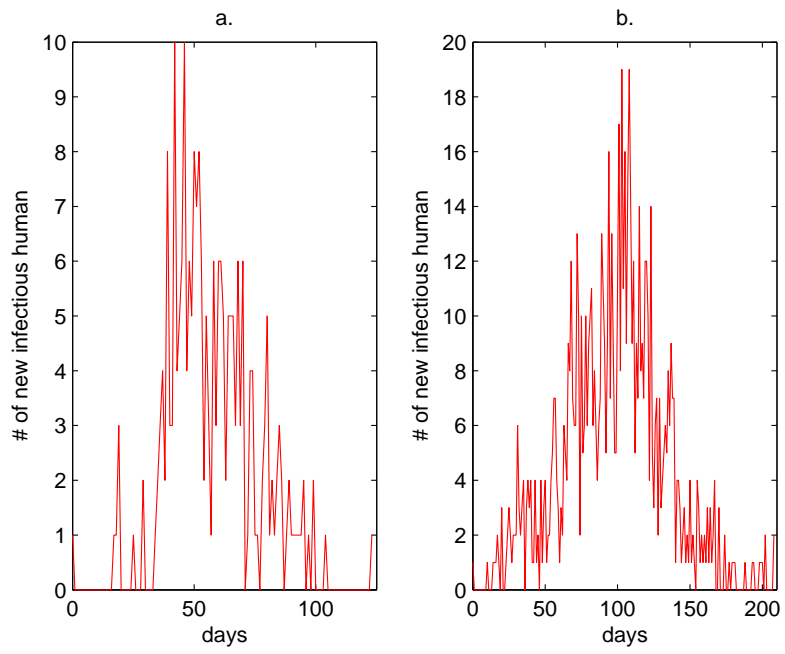
Figure 4: Number of new infectious humans per day in 2003 (a) and 2008 (b).
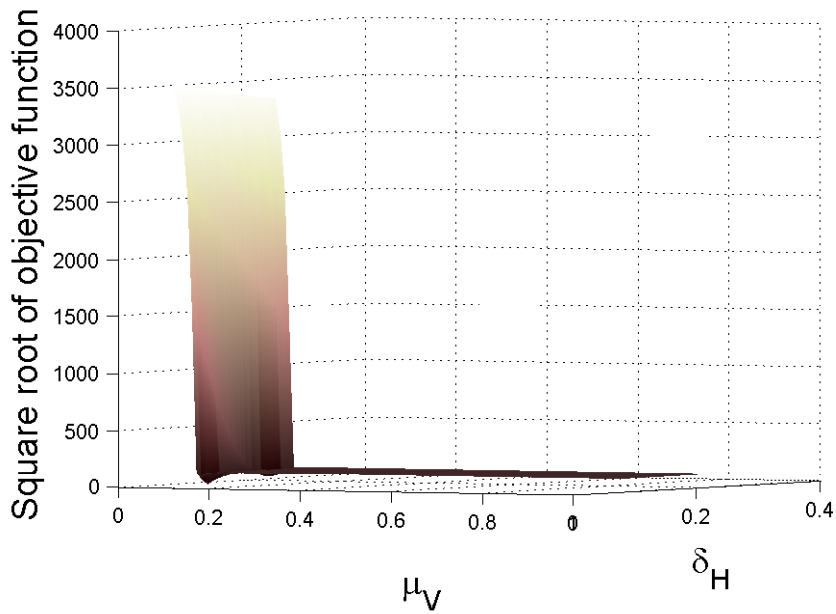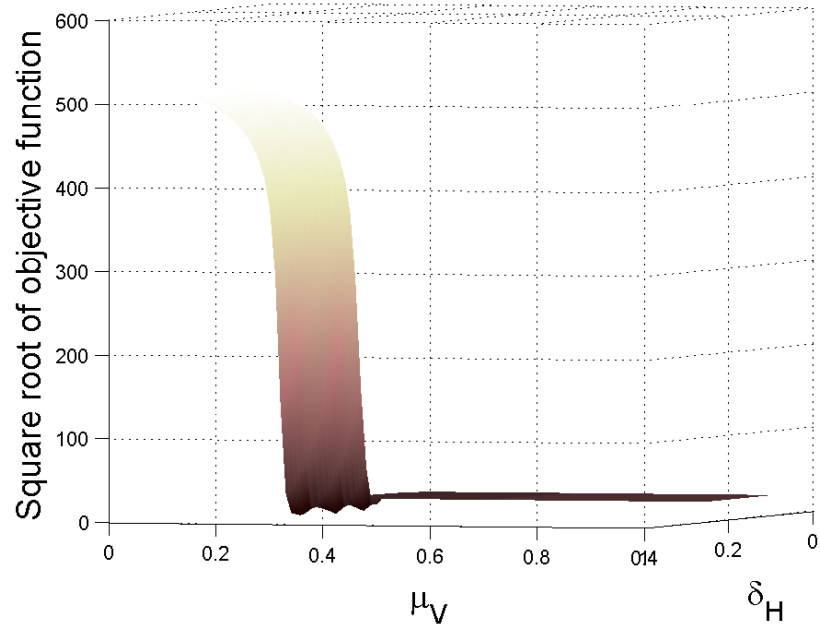
Figure 5: Examples of objective functions from scenarios II (a) and III (b) plotted as a function of (μ_V and δ_H) for fixed (θ_V and β_V)
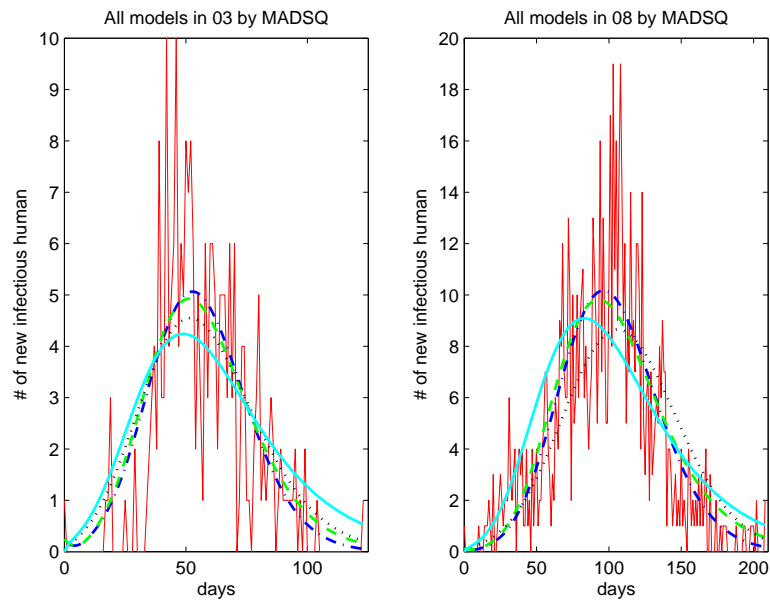
Figure 6: All Fitted results using MADSQ in two data sets: vSEIR (-.), fSEIR (−), vSIR (.), fSIR (-) and data (Red-)
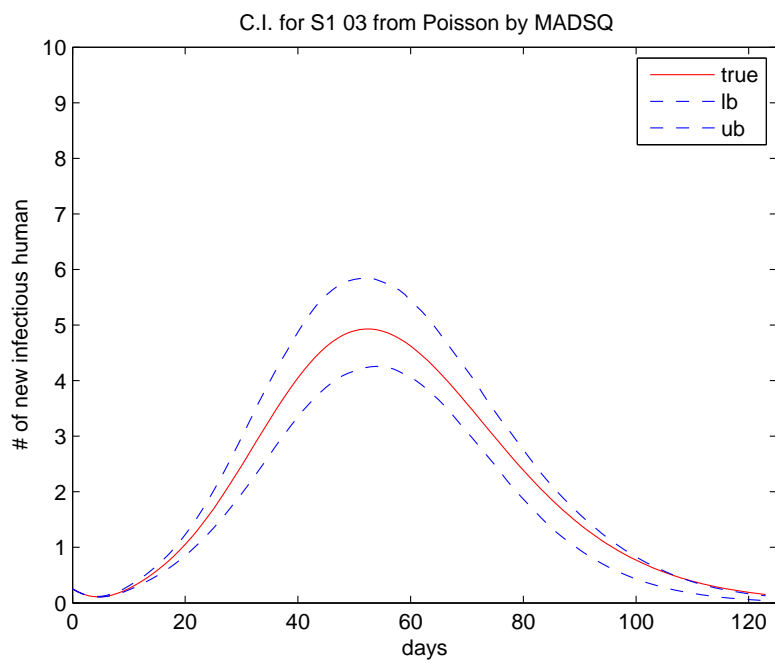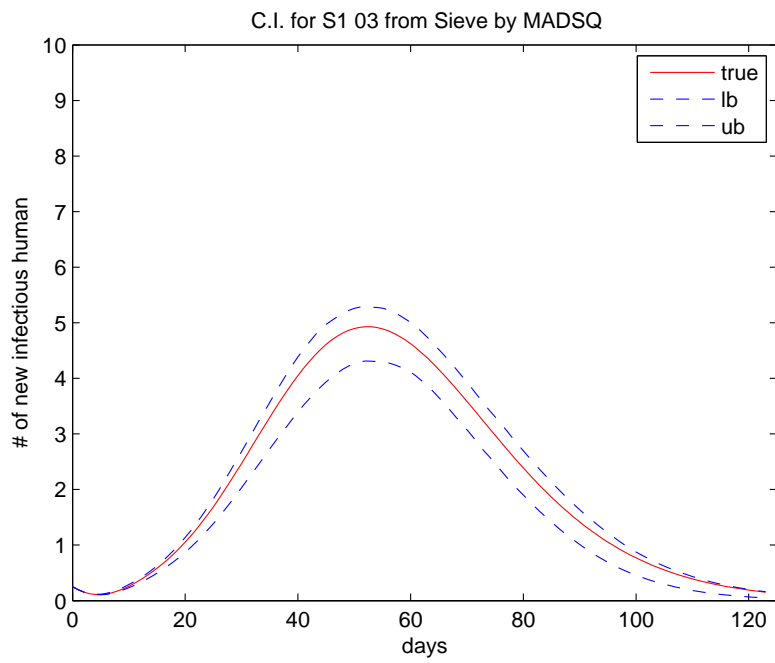
Figure 7: Bootstrap C.I. using MADSQ (–) in the 2003 data set by Sieve bootstrap and Parametric bootstrap with Poisson distribution

Table 2: Summary statistics for parameter estimates and M.C.S.E. for 200 simulated samples; variance=0.16; normal error; the bandwidth constant of the kernel smoothing ($\tilde{h} = ch^{5/9}$) is chosen to be 2 in all scenarios.

Scenario	Param	True	DTR Estimate	MCSE	MSE	MADSQ Estimate	MCSE	MSE
	θ_V	0.01	0.029	0.036		0.006	0.004	
I	β_V	0.32	0.336	0.026	0.089	0.321	0.008	0.001
	μ_V	0.08	0.096	0.029		0.079	0.003	
	δ_H	0.13	0.157	0.057		0.135	0.007	
Scenario	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.16	0.211	0.094		0.151	0.060	
II	β_V	0.54	0.581	0.083	0.158	0.576	0.062	0.051
	μ_V	0.23	0.279	0.089		0.217	0.054	
	δ_H	0.19	0.172	0.072		0.271	0.064	
Scenario	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.03	0.032	0.012		0.043	0.043	
III	β_V	0.14	0.146	0.013	0.613	0.159	0.013	3.474
	μ_V	0.06	0.064	0.010		0.075	0.039	
	δ_H	0.11	0.124	0.063		0.150	0.063	
Scenario	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.21	0.219	0.075		0.072	0.077	
IV	β_V	0.24	0.234	0.047	1.042	0.162	0.032	6.723
	μ_V	0.23	0.236	0.073		0.100	0.071	
	δ_H	0.22	0.195	0.061		0.148	0.063	

Table 3: Summary statistics for parameter estimates and M.C.S.E. for 200 simulated samples; variance=0.16; normal error; the bandwidth constant of the kernel smoothing ($\tilde{h} = \mathrm{c}h^{5/9}$) is chosen to be 2 using different initial points (Init) in scenarios II.

			DTR			MADSQ		
Parameter	*True*	Init I	Estimate	MCSE	MSE	Estimate	MCSE	MSE
θ_V	0.16	0.173	0.211	0.094		0.151	0.060	
β_V	0.54	0.592	0.581	0.083	0.158	0.576	0.062	0.051
μ_V	0.23	0.203	0.279	0.089		0.217	0.054	
δ_H	0.19	0.170	0.172	0.072		0.271	0.064	
Parameter	*True*	Init II	Estimate	MCSE	MSE	Estimate	MCSE	MSE
θ_V	0.16	0.142	0.126	0.089		0.182	0.003	
β_V	0.54	0.484	0.538	0.071	0.221	0.568	0.009	0.0001
μ_V	0.23	0.248	0.198	0.078		0.247	0.003	
δ_H	0.19	0.207	0.218	0.059		0.210	0.005	

Table 4: Summary statistics for model fitted parameter estimates in both data sets using different estimation algorithms in different scenarios.

| | a. 2003 Cairns outbreak | | | |
| | fSIR | | fSEIR A1 | |
Parameter	DTR	MADSQ	DTR	MADSQ
η_V	5.753×10^{-8}	0	0.031	0.001
β_V	0.100	0.118	0.305	0.335
μ_V	0.815	0.041	0.079	0.075
δ_H	0.227	0.326	0.080	0.166
	vSIR		vSEIR A1	
Parameter	DTR	MADSQ	DTR	MADSQ
θ_V	0.130	0.055	0.010	0.107
β_V	0.133	0.094	0.318	0.388
μ_V	0.157	0.089	0.082	0.157
δ_H	0.251	0.108	0.130	0.184
	b. 2008 Cairns outbreak			
	fSIR		fSEIR A1	
Parameter	DTR	MADSQ	DTR	MADSQ
η_V	1.17×10^{-17}	0.002	0.015	0.003
β_V	0.100	0.100	0.131	0.150
μ_V	0.026	0.027	0.040	0.037
δ_H	0.330	0.330	0.080	0.173
	vSIR		vSEIR A1	
Parameter	DTR	MADSQ	DTR	MADSQ
θ_V	0.133	0.127	0.169	0.050
β_V	0.101	0.096	0.217	0.139
μ_V	0.141	0.134	0.186	0.079
δ_H	0.249	0.250	0.231	0.090

3.0 VALIDATION OF A DENGUE AGENT-BASED MODEL BY COMPARING TO AN ODE MODEL, AND MATCHING WITH A REAL WORLD DATA SET

3.1 INTRODUCTION

Agent-based (AB) models have been developed to model the heterogeneous interaction relationships at an individual level under different conditions such as the container-inhabiting mosquito simulation model (CIMSiM) and the dengue simulation model (DENSiM) [Focks et al., 1995]. Given various epidemiological factors such as humans, mosquitoes, and environments from clinical studies [Watts et al., 1987, Kuno, 1995, Gubler, 1998], a study explored how dengue transmission dynamics are affected by the super-production phenomenon, when it is not able to be studied under the assumption of the homogeneous mixing between humans and mosquitoes [Padmanabha et al., 2012]. Recently, our group has been developing a Computational Arthropod Agents (CLARA) AB model. It represents both individual hosts (humans) and vectors (multiple life stages of *Ae. aegypti*: eggs, larvae, and the adults) via simulating the individual-level dynamics: the spread of the dengue virus and the effects of specific interventions. In addition, CLARA, which is spatially explicit, can track natural movement of both hosts and vectors as well as structural features (buildings), landscape (favorable and unfavorable zones), and individual oviposition sites. This framework allows us to explore spatial correlations in disease incidence and their dependence upon interventions. Nonetheless, model validation for such models can be challenging and subject to criticism due to lack of solid statistical theory.

Some previous studies have used ABM approaches [Deng et al., 2008, Perez and Dragicevic, 2009] but lack a clear validation procedure to build confidence in the modeling method-

ology representing the spatially and temporally heterogeneous spread of the disease. On the other hand, a study validated an AB model by comparing it with coarse epidemic curves [Jacintho et al., 2010]. In addition, a simplified human-only SEIR AB model and its corresponding ODE model were compared [Rahmandad and Sterman, 2008]. An interesting comparison result showed that the epidemic curve of the human infected cases in the AB model under the fully connected or the random move network circumstances is close to the curve in the ODE model because the fully connected or the random move networks in the AB model are similar to the ODE assumptions: homogeneity and perfect mixing within compartments.

Here, we propose a new ODE system with deterministic time lags and extend the idea in Rahmandad and Sterman [2008] to compare this ODE system with CLARA at a two-population model, since the interaction of different population is considered in both models. In addition, we propose a novel validation approach. After linking the spatial and temporal distances of the human infected cases together and forming a spatial-temporal cluster by the Laplacian matrix, we determine if there is a linear relationship between the spatial and temporal distances to a putative index case (PIdC) within a space-time cluster. The simulated data has been compared to the real world data in Cairns, Australia in 2003 [Vazquez-Prokopec et al., 2010]. These methods produce a robust approach for validating the CLARA model, which can provide insights that are easily applied to various parts of operations in the real world.

3.2 MODEL FORMULATION

3.2.1 Differential Equation Model

Direct comparison to our selected AB model requires a representation of coupled aquatic and adult stages in the mosquito population. Therefore, we developed a simplified continuous ODE model to reflect the interaction of humans with *Ae. aegypti* throughout the whole mosquito life cycle: from eggs to larvae to their adulthood [Bailey, 1975, Atkinson et al.,

2007]. An understanding of the life cycle of *Ae. aegypti* and how it interacts with humans is described before developing a dengue model.

Female *Ae. aegypti* lay eggs that develop into larvae during an emergence period with a relatively high survival probability. Larvae feed on microorganisms and particulate organic matter and change into pupae during an emergence period with a density-dependent survival probability. Pupae transform into adult *Ae. aegypti*. The entire aquatic life cycle lasts approximately 10-14 days until emerging into adult *Ae. aegypti*.

Female *Ae. aegypti* begin seeking blood after they have mated with male *Ae. aegypti*. If the target humans are infected, the dengue virus transmits from the humans to the female *Ae. aegypti*. Then the female *Ae. aegypti* become infectious in 10-14 days if they survive that long. On average, adult female *Ae. aegypti* can live about 10 days. Similarly, humans, infected by the infectious female *Ae. aegypti*, become infectious 4-7 days after exposure and remain infections for up to 12 days. For up to 6 months, they are not susceptible to other dengue virus infections (i.e. serotypes have cross-protection) and then become susceptible to the other serotypes of the dengue virus. Generally, humans can be infected with dengue fever twice during their lifespan although tertiary infections have been observed [Alvarez et al., 2006].

We make the following assumptions in our dengue ODE model

* Transition of larvae into adult *Ae. aegypti* is modeled directly without an intermediate pupae status.

* Only female *Ae. aegypti* are modeled while assuming there are enough males for successful mating with females.

* Only one-serotype of dengue virus is modeled (corresponding to a small outbreak).

* Infected female *Ae. aegypti* go through susceptible (S), exposed (E), and infectious (I) states (no recovered state and they remain at this state until death).

* Vector control interventions and seasonal effects of the outdoor household carrying-capacity of oviposition sites are not considered in the model.

* Symptomaticity is ignored in both the ODE model and the CLARA ABM for comparison.

* Humans go through susceptible (S), exposed (E), infectious (I), and recovered (R) states.

* The human mortality rate is ignored because of the short time spans in these simulations.

* The initial conditions of the ODE model are known.

The transition between various states can be represented using an SEIR compartmental model shown in Figure 8. Eggs and adult female mosquitoes die at a constant rate, but the larval mortality depends on the larval density [Legros et al., 2009]. Infected mosquitoes transmit the virus to susceptible humans, and infected humans transmit the virus to susceptible mosquitoes. Natural transition from S to E to I and to R in humans and from S to E to I in female *Ae. aegypti* occurs at various rates drawn from the literature. The model depicted in Figure 8 can be presented as a system of ODEs, which is similar to the model described in Chapter 1 [Atkinson et al., 2007] as follows:

$$N^V(t) = S^V(t) + E^V(t) + I^V(t)$$

$$\frac{dEg^V(t)}{dt} = \theta_V N^V(t) - \gamma_E Eg^V(t) - \mu_E Eg^V(t)$$

$$\frac{dL^V(t)}{dt} = \gamma_E Eg^V(t) - \gamma_L L^V(t) - \mu_L(L^V)L^V(t)$$

$$\frac{dS^V(t)}{dt} = \eta_V \gamma_L L^V(t) - \frac{\beta_V}{N^H} S^V(t)I^H(t) - \mu_V S^V(t)$$

$$\frac{dE^V(t)}{dt} = \frac{\beta_V}{N^H} S^V(t)I^H(t) - (1-P_V)^{1/\gamma_V} \frac{\beta_V}{N^H} S^V(t - \frac{1}{\gamma_V})I^H(t - \frac{1}{\gamma_V}) - \mu_V E^V(t)$$

$$\frac{dI^V(t)}{dt} = (1-P_V)^{1/\gamma_V} \frac{\beta_V}{N^H} S^V(t - \frac{1}{\gamma_V})I^H(t - \frac{1}{\gamma_V}) - \mu_V I^V(t)$$

$$N^H = S^H(t) + E^H(t) + I^H(t) + R^H(t)$$

$$\frac{dS^H(t)}{dt} = -\frac{\beta_H}{N^H} S^H(t)I^V(t)$$

$$\frac{dE^H(t)}{dt} = \frac{\beta_H}{N^H} S^H(t)I^V(t) - \frac{\beta_H}{N^H} S^H(t - \frac{1}{\gamma_H})I^V(t - \frac{1}{\gamma_H})$$

$$\frac{dI^H(t)}{dt} = \frac{\beta_H}{N^H} S^H(t - \frac{1}{\gamma_H})I^V(t - \frac{1}{\gamma_H}) - \delta_H I^H(t)$$

$$\frac{dR^H(t)}{dt} = \delta_H I^H(t)$$

where $Eg^V(t)$, $L^V(t)$, and $N^V(t)$ are the number of eggs, larvae, and the total number of female *Ae. aegypti*, respectively at time t. $S^V(t)$, $E^V(t)$, and $I^V(t)$ are the number of susceptible, exposed, and infectious female *Ae. aegypti*, respectively at time t. N^H is the total fixed number of human, which is calculated as a total of the number of susceptible, exposed, infectious, and recovered humans at time t, and are denoted by $S^H(t)$, $E^H(t)$, $I^H(t)$, and $R^H(t)$, respectively. Note that in the third equation, the death rate of larvae is usually modeled as $\mu_L(L^V) = \sigma(1 + L^V/\kappa)$, where κ is the carrying capacity of all oviposition sites. In addition, $(1 - P_V)^{1/\gamma_V} \frac{\beta_V}{N^H} S^V(t - \frac{1}{\gamma_V}) I^H(t - \frac{1}{\gamma_V})$ represents the number of new infected humans per day in the fourth and fifth equations, where P_V, γ_V, and β_V are the death probability of female *Ae. aegypti*, the female *Ae. aegypti* infection rate, and the virus transmission rate from humans to female *Ae. aegypti*, respectively. Although the factor $e^{-\mu_V/\gamma_V}$ is used in Atkinson et al. [2007], where μ_V is the death rate of female *Ae. aegypti*, this factor is replaced by $(1 - P_V)^{1/\gamma_V}$. That is because this ODE system is to compare with CLARA and the replaced factor more directly represents the survival probability of female *Ae. aegypti*. This replacement leads to increase the new infected female *Ae. aegypti* per day but decrease the exposed female *Ae. aegypti* per day.

There are 14 parameters including κ, σ, P_V, γ_V, β_V, and μ_V in the model in total. θ_V is the egg laying rate from the effective female *Ae. aegypti* (defined below). γ_E is the egg hatching rate. γ_L is the larvae eclose rate. η_V is the female to male *Ae. aegypti* ratio. γ_H and δ_H are the human infection and recovery rates. μ_E is the death rate of eggs. β_H is the virus transmission rate from female *Ae. aegypti* to humans. Note that μ_V is a function of P_V and it can be formulated as follows:

$$\mathbb{E}age_v = \frac{1}{\mu_V} = \sum_{i=1}^{n} i \frac{(1 - P_V)^i}{\sum_{i=1}^{n}(1 - P_V)^i}$$

where $\mathbb{E}age_v$ and n are expected and maximum ages of female *Ae. aegypti*, respectively.

Here, some standard parameters are substituted by a combined parameter because of the statistical identifiability. Those combined parameters are as follows:

* The egg laying rate from the effective female *Ae. aegypti* (θ_V) combines the number of progeny per female *Ae. aegypti* and the proportion of the effective female *Ae. aegypti*. Here an *effective* female *Ae. aegypti* means that a female *Ae. aegypti* is ready to lay eggs.

* The virus transmission rate from humans to female *Ae. aegypti* (β_V) combines the base biting rate and the probability that a bite infects a susceptible female *Ae. aegypti.*

* The virus transmission rate from female *Ae. aegypti* to humans (β_H) combines the base biting rate and the probability that a bite infects a susceptible human.

3.2.2 Experimental Design

For model comparisons, we link those parameters in CLARA to the ODE model based on the following procedures and parameter values from both models are in Table 1:

* The total carrying capacity (κ), σ, the female to male *Ae. aegypti* ratio (η_V), and the egg death rates (μ_E) are the same in both models.

* The extrinsic incubation period (T_e) of female *Ae. aegypti*, the intrinsic incubation period (T_i) and the disease duration (T_r) of humans are the inverse of the female *Ae. aegypti* infection rate (γ_V), the human infection rate (γ_H), and the recovery rate (δ_H), respectively.

* The egg hatched time (T_h), the larvae eclose time (T_l), and the averaged death age (T_d) of female *Ae. aegypti* are the inverse of the egg hatching rate (γ_E), the larvae eclose rate (γ_L), and the female *Ae. aegypti* death rates (μ_V), respectively.

* The multiplication of the number of progeny (N_e) per female *Ae. aegypti* and the proportion (P_e) of the effective female *Ae. aegypti* is the egg laying rate from the effective female *Ae. aegypti* (θ_V). Here the proportion of the effective female *Ae. aegypti* is a function of the genotrophic cycle length and other parameters related to the genotrophic cycle. It is a complicated relationship and we are not able to derive an explicit formula, but the empirical value is calculated directly in the AB model simulation.

* The multiplication of the base biting rate (B_r) and the probability (P_V) that a bite infects a susceptible female *Ae. aegypti* is the virus transmission rate from humans to female *Ae. aegypti* (β_V).

46

* The multiplication of the base biting rate (B_r) and the probability (P_H) that a bite infects a susceptible human is the virus transmission rate from female *Ae. aegypti* to humans (β_H).

In addition, we let the number of female *Ae. aegypti* be close at least in both models because female *Ae. aegypti* interact with humans directly.

Besides model comparisons, the model flexibility of CLARA can be performed by individual heterogeneity and network topology. In the previous study [Rahmandad and Sterman, 2008], in order to compare a simplified human-only SEIR AB model with a corresponding ordinary differential equation model, five different network structures of contacts among humans were created and the degree of human heterogeneity in the ABM could be varied. In our study, we varied the degree of heterogeneity in human and female *Ae. aegypti*, but only created two network structures of contacts among humans and female *Ae. aegypti* which are similar to the random move (RM) and small world (SW) in Rahmandad and Sterman [2008].

These two networks have important meaning and will be explained later. More detailed descriptions of the individual heterogeneity and the network topology are as follows. For heterogeneity among humans and female *Ae. aegypti* in CLARA, some parameters are allowed to vary in each individual. These are drawn from appropriate probability distributions. Two examples for each human are the intrinsic incubation period (IIP) and the disease duration. Female *Ae. aegypti* are more varied, examples being the extrinsic incubation period (EIP), the death rates of eggs and larvae, the genotrophic cycle length, and the factors related to the genotrophic cycle. In the heterogeneous condition, each human and female *Ae. aegypti* has their own personal settings for each parameter. Each parameter follows a certain distribution: IIP and EIP (Rayleigh distribution), the death rates of eggs and larvae (binomial), and others (uniform distribution between two certain values from previous clinical studies). In the homogeneous condition, the values of the parameter set in CLARA in each agent is identical to the values in the ODE model. However, the genotrophic cycle length and the factors related to the genotrophic cycle for each *Ae. aegypti* are varied in order to maintain the equilibrium states in the number of the eggs, larvae, and the female *Ae. aegypti*. Thus, both models are compared under the same equilibrium states.

For network structures of contacts among humans and female *Ae. aegypti*, RM is expected to create a scenario similar to the ODE model. Because the ODE model assumes homogeneity and perfect mixing within compartments, humans are allowed to be at any position at any time and the biting radius of the female *Ae. aegypti* is as large as the whole environment in RM. On the other hand, the SW, closer to the real world scenario, is used to compare with RM and SW means each human to be at their work place or school for two hours per day and at home for the rest of the day. Besides, the biting radius of the female *Ae. aegypti* is only a small number (R_b), roughly corresponding to the size of a house.

Moreover, two factors need to be controlled in order to compare different network topology. First, the number of female *Ae. aegypti* should be close at least in both network structures. The number of female *Ae. aegypti* is controlled by the number of larvae and it depends on the structure of the oviposition sites. If the locations of all oviposition sites are distributed in or on the houses (OVIP-In), the likelihood that female *Ae. aegypti* hatch eggs in the same oviposition sites will increase. Then the total number of larvae will decrease because the death rate of larvae is density dependent. However, if the locations of all oviposition sites are randomly distributed throughout the whole environment (OVIP-Random), the likelihood that female *Ae. aegypti* lay eggs in the same oviposition sites will decrease and the total number of larvae will increase. Therefore, under the distribution of OVIP-In, we compare SW with RM. In addition, under the distribution of OVIP-Random, RM is compared with the ODE model (more explanations of the realistic approach below).

Second, the mean number of links per node should be the same in both networks. The outbreak spreading depends on the mean number of links per node. If the mean number of links per node is larger, spreading is faster. Previously, a fixed number of real links is used to parameterize the model so that two networks have the same mean number of links per node [Rahmandad and Sterman, 2008]. Here, we considered a more flexible method to keep the mean number of links per node the same in both network topology. The details are described as follows:

The mean number of links per node in CLARA can be defined as the number of female *Ae. aegypti* multiplied by the base contact frequency and the percentage that the overall moving range of female *Ae. aegypti* and the human moving range overlap, divided by the

number of humans around those female *Ae. aegypti*. The overall moving range means the union of the biting and moving range of female *Ae. aegypti* in a certain area. First, let the mean number of links per node per day per property in RM and SW be denoted by $\bar{L}_r$ and $\bar{L}_s$, respectively; the base biting rate be denoted by B_r [Scott et al., 1993]; the total number of female *Ae. aegypti* in RM and SW be denoted by n_1 and n_2, respectively; the percentage that the overall moving range of female *Ae. aegypti* and the human moving range are overlapped in RM and SW be denoted by P_1 and P_2, respectively; the number of humans be the same in both networks and denoted by T; the number of houses be denoted by H (area surrounding a house); the number of the humans per property be t; the number of the female *Ae. aegypti* per property be n_{ij}, i = 1 for RM, 2 for SW, j from 1 to H. Then the mean number of links per node per day per property in RM ($\bar{L}_r$) are $\dfrac{n_1 B_r P_1}{T}$. Because humans are allowed to be at any position at any time and the biting radius of female *Ae. aegypti* is as large as the whole environment in RM, P_1 will equal to 1. In addition, the mean number of links per node per day per house in SW ($\bar{L}_s$) are $\dfrac{n_{2j} B_r P_2}{t}$, j from 1 to H. Because both female *Ae. aegypti* and humans are in the properties in SW, P_2 will equal to 1 as well. Thus, if the number of female *Ae. aegypti* per property are the same in SW and the total number of the female *Ae. aegypti* in both networks are the same, the mean number of links per node per day per property in both networks will be the same. It can be shown in the following formulas:

$$\bar{L}_s = \frac{n_2 B_r P_2}{Ht} = \frac{n_2 B_r P_2}{T} = \frac{n_2 B_r}{T} = \frac{n_1 B_r}{T} = \frac{n_1 B_r P_1}{T} = \bar{L}_r$$

$$n_2 = \sum_{j=1}^{H} n_{2j} = H n_{2j}, n_{2s} = n_{2t}, s,t \in \{1,\ldots,H\}$$

In order to let the number of female *Ae. aegypti* per property be the same in SW and because the number of female *Ae. aegypti* is proportional to the carrying-capacity of ovi-site under the same structure of the oviposition sites (OVIP-In), we let the carrying capacity of ovi-site be the same in each house and no ovi-site outside the house. Additionally, the number of ovi-sites and the carrying capacity of ovi-site are the same in both networks, so the total number of female *Ae. aegypti* in both networks will be the same. Of course, this is an ideal approach. In a more realistic approach, we would let all houses contain the majority of the

ovi-sites and some ovi-sites can be outside. On the other hand, in OVIP-Random, all houses contain 50% of the ovi-sites and the other 50% are randomly outdoor. In addition, household sizes and age structures are drawn from the census data, but the sizes of the workplaces and schools are varied. The number of ovi-sites in both structures are similar. Thus, the mean number of links per node per property in both networks will be close but not the same.

Under the above settings, we can see differences between the ODE model and RM in OVIP-Random are due to heterogeneity among individuals or the discrete and stochastic effect of individuals in CLARA. Differences between RM and SW in OVIP-In, however, are due to the network topology or the discrete and stochastic effect of individuals in CLARA.

3.2.3 New features in CLARA

The vector control interventions and the seasonality effect are in the real world data in Cairns, Australia in 2003 (the data is described in the section of Results) but the basic CLARA model did not have these features, so we added these features to the base model first. In order to test the linear relationship between the spatial and temporal distances to a PIdC within a space-time cluster, we implemented the clustering algorithm as described in Vazquez-Prokopec et al. [2010] as well.

3.2.3.1 Vector Control Intervention Algorithm We developed an intervention kernel using algorithms that match real world control strategies for reducing mosquitoes, larva, and containers. First, the interventions are scheduled either when a person is infected by the dengue virus and the fever is symptomatic, or when a small probability random event occurs. Then, at a scheduled date, the interventions are carried out within a certain distance of the house where the infected person lives. If there is any household member who does not consent to the indoor intervention, the intervention will only be implemented in a certain range outside of their house. In addition, the differences between the algorithms among these three interventions are as follows: the mosquitoes and larva which are sprayed by the insecticide are assigned a higher death rate, and the rate follows the distribution that has a fixed value before a certain date and decreases exponentially after that certain date.

However, a certain percentage of the eggs and larva is removed from the oviposition sites in the container-reduced intervention.

3.2.3.2 Seasonality Algorithm When considering that the outdoor household carrying capacity of oviposition sites is a seasonal function of time, we let the outdoor household carrying capacity of oviposition sites (K_{out}) be the mean outdoor carrying capacity of oviposition sites (K_{mean}) times a cosine function whose maximum value is at the day (D_{hr}) with the highest rainfall precipitation at that certain location in a given year and periodic cycle length (D_{year}) is a year. The formula is as follows:

$$K_{out}(t) = K_{mean}(1 + cos(2\pi(t - D_{hr})/D_{year}))$$

3.2.3.3 Clustering Algorithm In order to form a space-time cluster in a simulation, the Laplacian clustering matrix is first formed based on time and distance criteria given the simulated data of the infected humans at a certain time and location. The time and distance criteria are the total incubation periods and the mosquito dispersal distance, respectively. Each row of the Laplacian matrix has the number 1 or 0 to represent the relationship between these infected humans. 1 means the two assigned infected humans are in the same cluster and 0 means the two assigned infected humans are not in the same cluster. Even if two assigned infected humans in the same row are not in the same cluster, they will be collected into the same cluster if there are other mutually infected humans in these two rows of the Laplacian matrix. Then, all infected humans are grouped into distinct clusters and the spatial and temporal distances to a PIdC within a space-time cluster is plotted.

3.3 RESULTS

3.3.1 Model Comparison

Key parameter values from both models are in Table 5 and the procedures to link them are described in the method section above. In addition, total human population is 502, initial

female *Ae. aegypti* population is 2 times larger than total human population [Williams et al., 2013], the simulation time period is 180 days, and all AB model results in the Figures are an average of 100 simulations. Since the time unit in CLARA and the ODE model is an hour and a day, respectively, the results in CLARA are rescaled to match the time unit in the ODE model.

In this study, different models or networks are compared by exploring three standard measures of the SEIR model [Rahmandad and Sterman, 2008]: the maximum symptomatic infected population (*peak prevalence*, I_{max}), the time from initial exposure to the maximum of the symptomatic infected population (the *peak time*, T_p), and the fraction of the population ultimately infected (the *final size*, F).

The differences between the ODE model and RM in the OVIP-Random structure for different state variables in the heterogeneous/homogeneous condition are shown in Figure 9. The number of the total female *Ae. aegypti* are similar in both models (Figure 9.c-e). The averaged epidemic grows faster in the ODE model (Figure 9.e and 9.h: $T_p = 55$ versus 81/82 days in the female *Ae. aegypti* and 39 versus 67/67 days in humans). Both models have similar averaged peak prevalence (Figure 9.e and 9.h: $I_{max} = 701$ versus 1,106/1,061 female *Ae. aegypti* and 114 versus 130/117 humans). In both models, all humans are ultimately infected. Unfortunately, the number of eggs and larvae of the *Ae. aegypti* are not able to be matched in both models (Figure 9.a-b). Given the same egg laying rate from the effective female *Ae. aegypti* in both models, although the number of the eggs and larvae of the *Ae. aegypti* are close (shown in Figure S1.a-b), the number of the female *Ae. aegypti* will be more in the ODE model and the rest of the results will differ in both models (Figure S1.c-h). However, it is more important to have the same magnitude in the total number of the female *Ae. aegypti* for model comparison, because this directly affects transmission. Thus, the egg laying rate from the effective female *Ae. aegypti* is rescaled from 1.3 to 0.61 per day in the ODE model.

The differences of RM in both structures of the oviposition sites for different state variables in the heterogeneous condition are shown in Figure 10. Since the likelihood that female *Ae. aegypti* hatch eggs in the same oviposition sites in both structures is different, the number of the eggs, larvae, and the total female *Ae. aegypti* are more in the OVIP-Random

structure (Figure 10.a-e). The averaged epidemic grows faster in the OVIP-Random struc-
ture (Figure 10.e and 10.h: $T_p = 81$ versus 109 days in the female *Ae. aegypti* and 67 versus
95 days in humans), and has a larger averaged peak prevalence (Figure 10.e and 10.h: I_{max}
= 1106 versus 121 female *Ae. aegypti* and 130 versus 53 humans), as well as finally infecting
more humans ($F = 100\%$ versus 98%).

The differences between RM and SW in the OVIP-In structure for different state variables
in the heterogeneous condition are shown in Figure 11. When the number of the eggs, larvae,
and the total female *Ae. aegypti* are similar in both networks (Figure 11.a-e), the averaged
epidemic grows faster, has larger averaged peak prevalence, and finally infects more humans
in the RM network ($F = 98\%$ versus 40%). However, the peak time of SW is unobtainable
within the duration of this simulation.

In addition, we show some results for alternative scenarios. The difference between RM
in the OVIP-In structure in the heterogeneous and the homogeneous condition are shown in
Figure S2. When the number of the eggs, larvae, and the total female *Ae. aegypti* are similar
in both conditions (Figure S2.a-e), the averaged epidemic grows similarly (Figure S2.e and
S2.h: $T_p = 109$ versus 110 days in the female *Ae. aegypti* and 95 versus 96 days in humans),
and both conditions ultimately have similar averaged peak prevalence (Figure S2.e and S2.h:
$I_{max} = 121$ versus 121 female *Ae. aegypti* and 53 versus 52 humans) and infect humans at
a same level ($F = 98\%$ versus 98%). When comparing the difference between SW in the
OVIP-In structure in the heterogeneous and the homogeneous condition, similar patterns
for both conditions are shown in Figure S3. However, the peak time of SW is unobtainable
within the duration of this simulation.

Moreover, for the sensitivity analysis of the population scale, we create two more scenarios
for 800 and 210 human populations in 180 and 118 simulation days, respectively [Rahmandad
and Sterman, 2008]. The differences between the ODE model and two ABM networks in
the heterogeneous condition for the 800 human population scenario are shown in Figure
S4. The results are similar to those in the 502 human population scenario (Figure 9-10-11).
When the number of the total female *Ae. aegypti* are similar in the ODE model and RM
in OVIP-Random (Figure S4.c-e), the averaged epidemic grows faster in the ODE model
(Figure S4.e and S4.h: $T_p = 58$ versus 83 days in the female *Ae. aegypti* and 42 versus 68

days in humans). RM in OVIP-Random has a larger averaged peak prevalence of female *Ae. aegypti* but its averaged peak prevalence is similar to the ODE model's averaged peak prevalence in humans (Figure S4.e and S4.h: I_{max} = 1,020 versus 1,903 female *Ae. aegypti* and 163 versus 221 humans). Finally, both models infect all humans. On the other hand, when the number of the total female *Ae. aegypti* are similar in RM and SW in OVIP-In (Figure S4.c-e), the averaged epidemic grows faster in the RM structure. RM in OVIP-In has larger averaged peak prevalence. Finally, more humans are infected in the RM network (F = 98% versus 42%). However, the differences between RM and SW are very large and the peak time of SW is unobtainable within the duration of this simulation because of the large simulation environment. When comparing the differences between the ODE model and two ABM networks in the heterogeneous condition for the 210 human population scenario, patterns (Figure S5) are similar to those of the 800 human population (Figure S4).

3.3.2 Data Validation

The data set is from one of the largest outbreaks that occurred in the city of Cairns, Australia in 2003. The onset of symptoms in the first reported case occurred on January 22 in Parramatta Park (PP) but the dengue spread was not identified until March 2. Mosquito control measures were initiated the next day. A total of 383 laboratory-confirmed mild DENV-2 symptomatic cases were registered within urban Cairns over the 25-week epidemic period. There is a linear relationship between the spatial and temporal distances to a PIdC within a space-time cluster in Cairns (Figure 12) [Vazquez-Prokopec et al., 2010], and the detailed data description can be found in Vazquez-Prokopec et al. [2010] as well.

   In order to see if this observation of the linear relationship can be replicated in CLARA, some initial settings for those new features in CLARA are listed:

 * The time and space criteria of the clustering algorithm are 20 days and 100 meters, respectively.

 * The first day of the control interventions is the 42nd day after the date that the first human was infected.

 * The intervention coverage radius is 100 meters.

* 80% of *Ae. aegypti* within 100 meters will be sprayed.

* The mortality rate of larvae and *Ae. aegypti* will be 0.9 per day before the 21st day and have a 0.083 exponential decay after the 21st day of each vector and larvae control intervention.

* 60% of the eggs and larvae will be removed under each container control intervention.

* The intervention compliance ratio of the infected household is 40%.

* The symptomatic ratio of the symptomatic to inapparent infections is 25%.

* The time region to schedule the interventions is from the 1st day to the 25th day after the infection or from the 5th day to the 25th day after the infection for either the symptomatic cases or the random event cases.

* The maximum value of the seasonality function is the 55th day after the date that the first human was infected.

* Total realizations are 10 with the human population 2,828.

When using the clustering algorithm, we require the number of the elements in each cluster be more than 2. We can see that there is a linear relationship between the spatial and temporal distances to a PIdC within a space-time cluster from one of the CLARA realizations in Figure 12.b. In addition, we tried to fit a linear regression line to the data of the spatial and temporal distances to a PIdC within each space-time cluster in the real world data and the CLARA realizations. The mean and the standard deviation of the slope (measured in meters per week) of the line in the real world data and CLARA are (25.87, 24.07) and (25.5033, 32.9605), respectively. From another point of view (Figure 12.c-d), except there are some outliers of the slopes in each realization, most of the slopes in CLARA realizations are similar to the ones in the real world data.

3.4 DISCUSSION

Computational modelers always face trade-offs. ODE models are the standard to model dynamic systems involving interactions between various populations because of their solid statistical theory and quick calculations. The assumptions of the homogeneity and perfect mixing of the ODE model may not accurately represent the real world. On the other hand, AB models may lack the solid statistical theory, but can model heterogeneity at the individual level. Here, we propose a set of ODEs with deterministic time lags to model the incidence of dengue virus infection and compare it with our CLARA AB model. In CLARA AB model, we compare different network topology as well. The results not only show similarity in both models, but also indicate the flexibility of CLARA.

Differences between the ODE model and CLARA are due to the heterogeneity among individuals or the discrete and stochastic effect of individuals in CLARA. In three standard public health measures, the peak time resulting from the ODEs is earlier than that in the AB model networks, but the mean peak time in RM is closer to the ODE model. This establishes the correspondence of ODE and AB models in the fully randomized case, confirming the flexibility of AB model to explore meaningful differences between heterogeneous and homogeneous dynamics. In addition, the results of the human peak prevalence and the final size in the ODE model are similar to RM, but have huge differences compared to SW in heterogeneous or homogeneous conditions (in Figure 9-10-11 and S4-S5). Moreover, there are only small differences in both conditions no matter what the structures of the network or the oviposition sites are (in Figure 9, S2, and S3).

For validation purposes, we took a different approach than comparing the coarse epidemic curve to the real world data because of the high asymptomaticity of incidences and the limited data availability. The ratio of symptomatic to inapparent infections varies from one reference to another [Yoon et al., 2012, Endy et al., 2011]. Even if a specific ratio can be calibrated by one data set, the ratio may not be able to be validated from another similar data set. Thus, we compare the simulated pattern to the pattern in the real world data and a linear relationship between the spatial and temporal distances to a PIdC within a space-time cluster which is shown in the real world data can be reproduced from CLARA in a small

population scale. Although the linear relationship is replicated under the assumption of 25% symptomatic ratio of symptomatic to inapparent infected humans, the relationship can be shown no matter what value of symptomatic ratio is specified.

Unfortunately, the number of eggs and larvae of the *Ae. aegypti* are not able to be matched in both models in the original settings, so the egg laying rate from the effective female *Ae. aegypti* needs to be rescaled to a smaller value in the ODE model to match the number of the female *Ae. aegypti*. We still observe some differences between a homogeneous and perfectly mixed scenario and the current CLARA. However, these results demonstrate preliminary validation of CLARA at a certain level and provide us with the confidence that our simulation results correspond to a well constructed ODE model and to the real world.

3.5 FIGURES AND TABLES
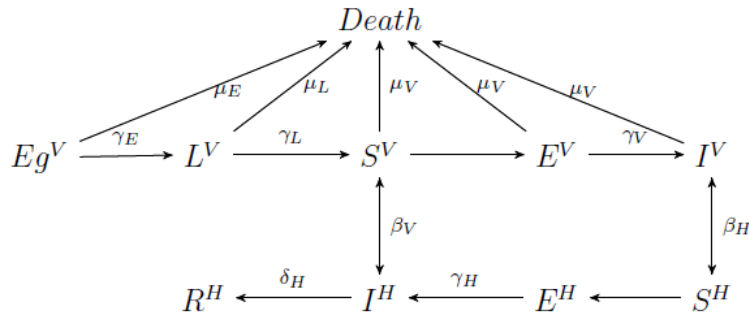
Figure 8: The diagram of the simplifed model of the interaction between humans and mosquito in different stages
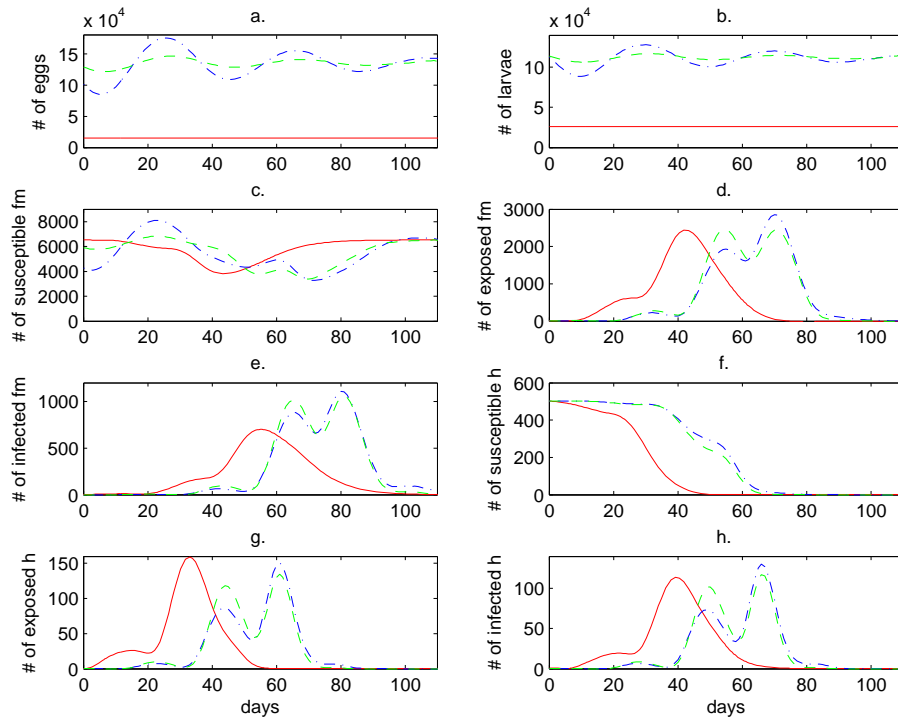
Figure 9: Model comparison between ODE (Red-) and heterogeneous (Blue-.) or homogeneous (Green–) RM network ABM for different state variables when the number of total female *Ae. aegypti* are at the same level in both models
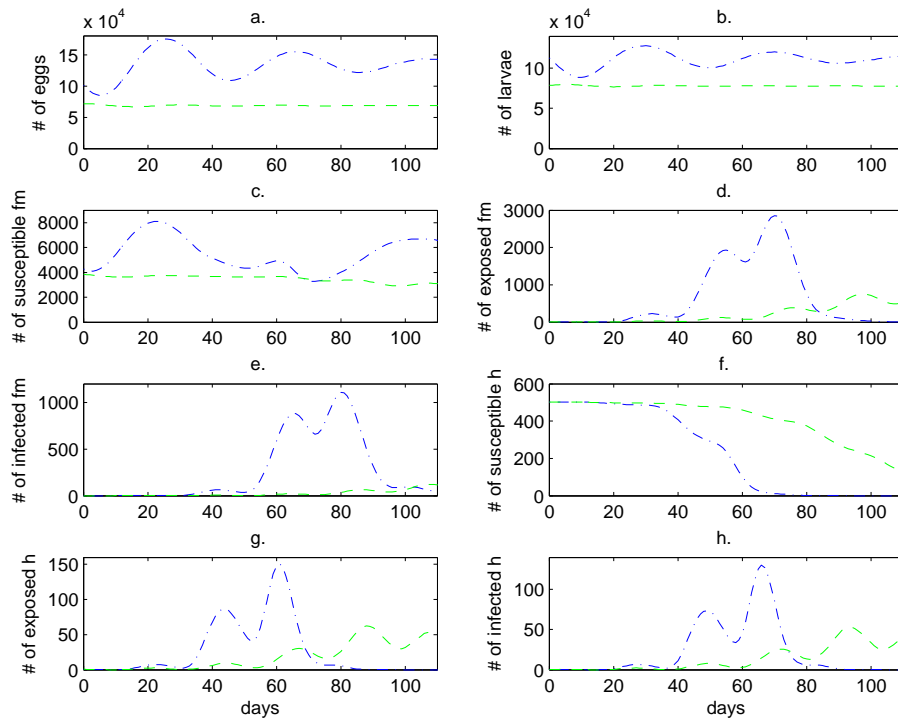
Figure 10: Heterogeneous RM network ABM in OVIP-Random (Blue-.) and OVIP-In (Green–) for different state variables

59
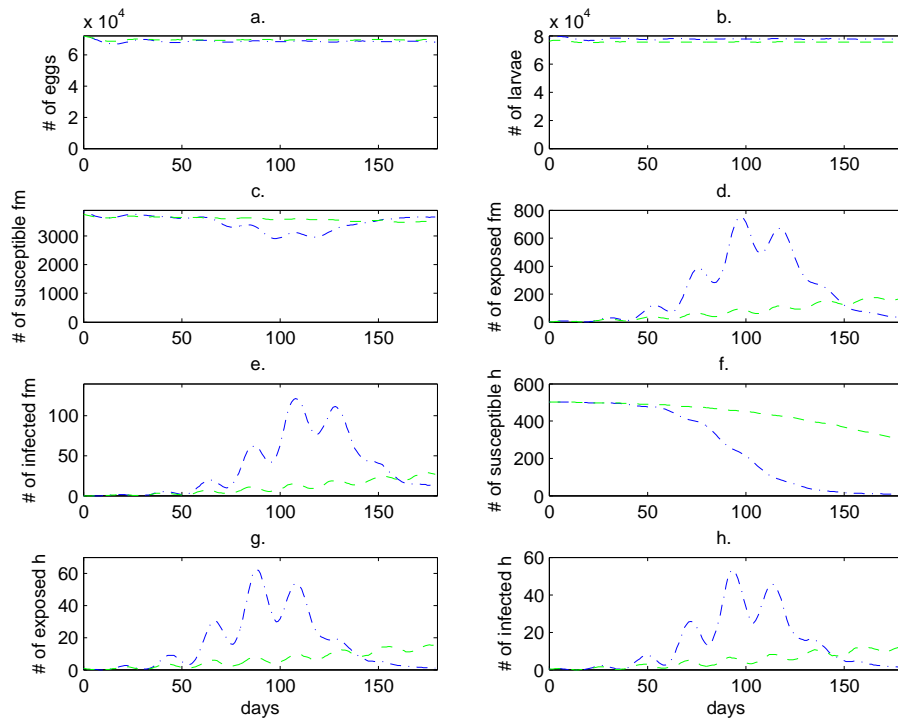
Figure 11: Heterogeneous RM network ABM (Blue-.) and SW network ABM (Green–) for different state variables

Figure 12: The linear relationship between the spatial and temporal distances to a putative index case (PIdC) within a space-time cluster in Cairns (left) and CLARA (right)

Table 5: Parameters in both models: ODE (above) and CLARA (below)

Parameter	Value	Units	Parameter	Value	Units
σ	0.025		θ_V	0.61	1/Day
η_V	0.5		γ_E	0.25	1/Day
κ	13000		γ_L	0.07	1/Day
β_H	0.51	1/Day	μ_E	0.01	1/Day
γ_H	0.174	1/Day	β_V	0.51	1/Day
δ_H	0.207	1/Day	μ_V	0.141	1/Day
			γ_V	0.086	1/Day
Parameter	Value	Units	Parameter	Value	Units
σ	0.025		N_e	8	egg
η_V	0.5		P_e	0.076	
κ	13000		T_h	4	Day
B_r	0.567	1/Day	T_l	14	Day
P_H	0.9		μ_E	0.01	1/Day
T_i	5.75	Day	P_V	0.11	
T_r	4.831	Day	T_d	7.1	Day
R_b	5	Meter	T_e	11.6	Day

4.0 CONCLUSIONS AND DISCUSSION

4.1 SUMMARY

In the first part of this dissertation (Chapter 2), we proposed a new ODE model (vSEIR). The proposed model, unlike the simplified SIR model, includes the exposed state variables and the varied birth rate of the female *Ae. aegypti*. Additionally, it does not require the information about the aquatic stages of the *Ae. aegypti*. We demonstrated that the proposed vSEIR ODE model provides a better fit to the data than the other three existing ODE models. We also compared two discretization methods for initial value problems: a derivative-free mesh adaptive direct search method with quadratic models (MADSQ) and a derivative trust region (DTR) method.

Our simulation studies showed that MADSQ can provide a better solution to the ODE compared to DTR when the parameter space has many local minima. Further simulation results for parameter estimation in different scenarios are included in Appendix B. Table S1, S2, S3, and S4 summarize the performance of the parameter estimation in both methods when the simulated data variance (Z_t) ranges between 0.1 to 2.0 for scenarios I to IV. The results showed that approximately unbiased estimators may be obtained in scenario I by MADSQ (Table S1) and in scenario IV by DTR (Table S4) even if the simulated data variance (Z_t) is large. The MADSQ estimates are poor in scenario IV even when the simulated data variance (Z_t) was small. However, approximately unbiased estimators may be obtained by both methods in scenario presented in Table S5 which follows the same assumption as in scenario IV even if the simulated data variance (Z_t) was large. Table S6 displays the estimator and the Monte Carlo standard error (MCSE) of the virus transmission rate from humans to female *Ae. aegypti* (β_V) keeping all others fixed using only the simulated data

before the date of the vector control intervention when the simulated data variance (Z_t) equals to 2.0 for scenarios I to IV. The results showed that the unbiased estimator would be achieved in scenarios I and II, which include 40 time points, but only approximately unbiased estimator are obtained in scenarios III and IV, since these two scenarios only have 22 time points. However, as the simulated data variance get small, unbiased estimators tend to be obtained as shown in Table S7.

In the second part of this dissertation, we proposed two methods to validate a dengue **C**omputationa**L AR**thropod **A**gents (CLARA) AB model. First, we compared our CLARA AB model with a proposed ODE model with deterministic time lags. The results showed the similarity of the two models. Second, because the ratio of symptomatic to inapparent infections varies from one reference to another [Yoon et al., 2012, Endy et al., 2011] and the lack of existing data, we avoided comparing the epidemic curve to the real world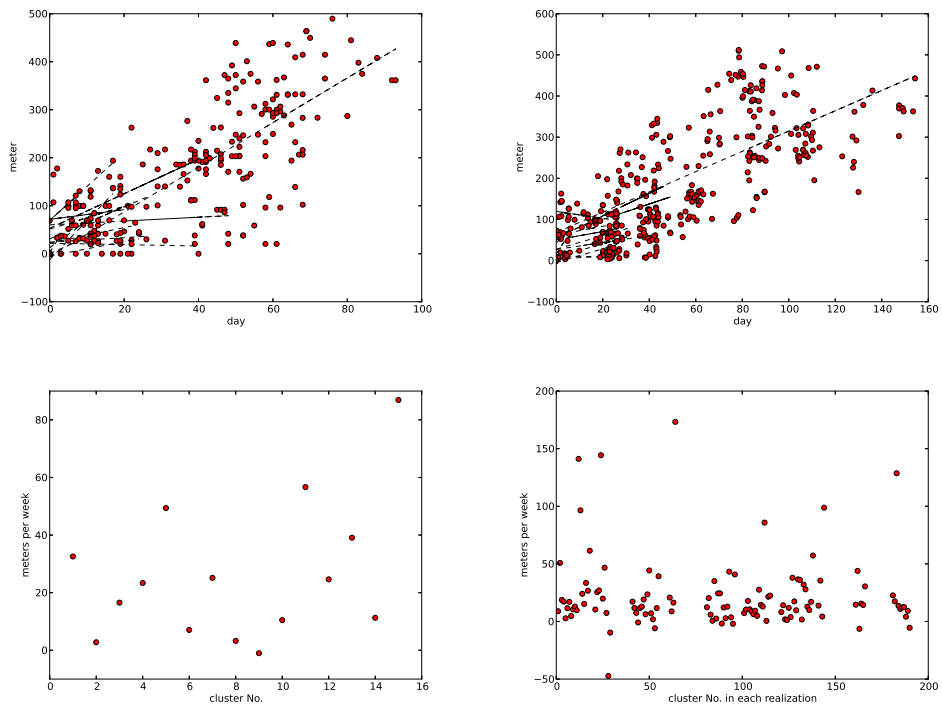 data, but matched the simulated data to the real world data by a documented linear relationship between the spatial and temporal distances to a putative index case (PIdC) within a space-time cluster. The results showed that the linear relationship can be replicated.

4.2   FUTURE WORK

4.2.1   A Proposed Dengue ODE Model with Asymptomatic Compartment

In Chapter 2 we assumed that all infected humans are symptomatic. We would like to extend the model to consider the fact that a fraction of infected humans may be asymptomatic in reality. In addition, we may assume that only a portion of the asymptomatically infected humans can transmit the virus and assume that the virus transmission rate from female *Ae. aegypti* to humans and vice versa are time dependent [Reiner Jr. et al., 2014]. After following the same assumptions in Chapter 2 except that not all infected humans are symptomatic, the vSEAIR model can be presented as an ODE model as follows:

$$N^V(t) = S^V(t) + E^V(t) + I^V(t)$$

$$\frac{dS^V(t)}{dt} = \theta_V N^V(t) - \frac{\beta_V(t)}{N^H(t)} S^V(t)(\alpha_H A^H(t) + I^H(t)) - \mu_V S^V(t)$$

$$\frac{dE^V(t)}{dt} = \frac{\beta_V(t)}{N^H(t)} S^V(t)(\alpha_H A^H(t) + I^H(t)) - \gamma_V E^V(t) - \mu_V E^V(t)$$

$$\frac{dI^V(t)}{dt} = \gamma_V E^V(t) - \mu_V I^V(t)$$

$$N^H(t) = S^H(t) + E^H(t) + A^H(t) + I^H(t) + R^H(t)$$

$$\frac{dS^H(t)}{dt} = -\frac{\beta_H(t)}{N^H(t)} S^H(t) I^V(t)$$

$$\frac{dE^H(t)}{dt} = \frac{\beta_H(t)}{N^H(t)} S^H(t) I^V(t) - \gamma_H E^H(t)$$

$$\frac{dA^H(t)}{dt} = \gamma_H(1 - \kappa_H) E^H(t) - \delta_H A^H(t)$$

$$\frac{dI^H(t)}{dt} = \gamma_H \kappa_H E^H(t) - \delta_H I^H(t)$$

$$\frac{dR^H(t)}{dt} = \delta_H(A^H(t) + I^H(t))$$

where $N^V(t)$ is the number of the total number of female *Ae. aegypti* at time t and $S^V(t)$, $E^V(t)$, and $I^V(t)$ are the number of susceptible, exposed, and infectious female *Ae. aegypti*, respectively at time t; Additionally, N^H is the total fixed number of humans at time t and is calculated as a total of the number of susceptible, exposed, infectious symptomatically, infectious asymptomatically, and recovered humans at time t denoted by $S^H(t)$, $E^H(t)$, $I^H(t)$, $A^H(t)$, and $R^H(t)$, respectively.

There are nine parameters in the vSEAIR model mentioned above. θ_V is the birth rate of *Ae. aegypti* in the model with the varied birth of the female *Ae. aegypti*, κ_H is the proportion of the symptomatically infected humans, and α_H is the proportion of the asymptomatically infected humans which can transmit the virus. γ_V (μ_V) is female *Ae. aegypti* infection (death) rate. γ_H (δ_H) is the human infection (recovery) rate. The virus transmission rates from humans to female *Ae. aegypti* and from female *Ae. aegypti* to humans are time-dependent and are denoted by $\beta_V(t)$ and $\beta_H(t)$, respectively.

Compared to the vSEIR model, this model involves three non-identifiable parameters: α_H, γ_H, and κ_H. This issue needs to be resolved before fitting the model to the data. One way to address it is to assume one of the parameters to be known. For example, γ_H is easier to be determined in the literature but κ_H has larger variability [Yoon et al., 2012, Endy et al., 2011].

4.2.2 Other Estimation methods in Ordinary Differential Equation Models

In Chapter 2, we considered estimating parameters in a simplified dengue ODE model by using the derivative free MADSQ method. We demonstrated via simulations that the derivative free MADSQ method is better than derivative trust region (DTR) when the parameter space has many local minima. However, we would like to confirm this by comparing the estimation results with another derivative estimation method, interior-point filter line-search algorithm [Wächter and Biegler, 2006]. There are two iterative approaches to find parameter estimates: trust region and line search. When using the trust region approach, DTR first chooses a step size (the size of the trust region) and then a step direction. The interior-point filter line-search follows an inverse mechanism: choose a step direction and then a step size. When seasonality such as recruitment, mortality and biting rates, and duration of EIP of the dengue ODE model is included in the model [Bartley et al., 2002], the interior-point filter line-search algorithm can perform well [Word et al., 2012].

4.2.3 Dengue Stochastic Differential Equation (SDE) Models and Parameter Estimation

While parameters can be estimated well by minimizing an objective function via least squares for nonlinear deterministic ODE system in Chapter 2, the stochastic nature of the problem is usually unaccounted for. Thus, we would like to develop and implement a nonlinear stochastic differential equation system (SDE) for dengue dynamics. Three main approaches have been proposed in the literature. The first approach is to estimate a numerical solution set of the SDE system directly. But it is difficult to have an accurate numerical solution set of the SDE system by using an objective function based on least squares with the discretization

methods for initial value. Solutions approximated by higher order integration schemes are hardly tractable; solutions approximated by lower order integration schemes such as Euler are biased unless the time scales of integrating and sampling are equal and extremely small, such as $0.0001s$ [Timmer, 2000]. The other two approaches are likelihood-based. The first likelihood-based approach is to use simulated maximum likelihood estimation. This is based on the fact that the likelihood function for a sampled time series of length N with the Markov property can be written as

$$L(x(t_1), x(t_2), \ldots, x(t_N); \theta) = \pi(x(t_1))\Pi_{i=1}^{N-1}p(x(t_{i+1})|x(t_i)), \theta).$$

Then the conditional densities $p(x(t_{i+1})|x(t_i))$ can be estimated by methods such as kernel estimation. While this approach works in univariate models [Singer, 2002], it is not applicable in multivariate models due to its high computational burden. The second likelihood-based approach is to use the Bayesian imputation or approximated function methods. The Markov chain Monte Carlo (MCMC) scheme is used in the Bayesian imputation method. This strategy works well in univariate models [Elerian et al., 2001], as well as in multivariate models with partial state variables and observed errors [Golightly and Wilkinson, 2008]. On the other hand, the extended Kalman filter (EKF) scheme offers an iterative procedure to approximate the likelihood function. This strategy works efficiently in univariate models [Singer, 2002]. It can also be used to approximate the posterior distribution of the parameters and a reasonable set of estimators can be obtained from it [Mbalawata and Särkkä, 2013]. Therefore, we would like to see if the Bayesian imputation or approximated function methods can be applied to our model and used to develop an accurate solution, as it is computationally more efficient than other methods.

4.3   PUBLIC HEALTH SIGNIFICANCE

As mentioned in Chapter 1, dengue fever is currently the world's fastest growing vector-borne disease. While more than 2.5 billion people living in areas of risk, around 50-100 million people are infected every year, mostly in urban and semi-urban areas [Halstead,

2007]. Although countries have implemented different vector control strategies [Yeap et al., 2011], these interventions have achieved only limited success, and the trend of the global spread is expanding [Horstick et al., 2010]. No licensed dengue vaccine is available for the public either, although several vaccine candidates are currently being evaluated in clinical studies [Guy et al., 2011]. Thus, global control of dengue fever is a major public health problem with significant economic, political, and social impact [Kyle and Harris, 2008] and developing different models to tackle the problem is the main task in this field. Modeling the spread of dengue, however, is challenging not only because it involves numerous complex factors such as the interactions between humans and mosquitoes, multiple coexisting virus serotypes, and high asymptomaticity of the initial infection [Kyle and Harris, 2008], but also because of the lack of existing data [Andraud et al., 2012].

An efficient dengue surveillance system can provide decision makers with a reasonable solution to existing and potential public health problems preventing the speed of dengue during an outbreak. Accuracy in modeling of dengue incidence over time is an important first step to develop such a system.

In this dissertation, we proposed and validated several methods of modeling that can be easily used in practice. This will provide better understanding of the mechanisms of dengue outbreaks and hence allow public health professionals to develop proper interventions.

PROOF OF MATHEMATICAL PROPERTIES OF VSEIR MODEL

A.1 PROOF OF MATHEMATICAL PROPERTIES

PROOF OF THEOREM 3 To prove theorem 3, the Jacobian of the system (1) at equilibrium points x_{dfe} is given by:

$$J(x_{dfe}) = \begin{pmatrix} -\gamma_V - \mu_V & 0 & 0 & \frac{\beta_V N^V}{N^H} & 0 \\ \gamma_V & -\mu_V & 0 & 0 & 0 \\ 0 & \frac{\beta_H(N^H - R^H)}{N^H} & -\gamma_H & 0 & 0 \\ 0 & 0 & \gamma_H & -\delta_H & 0 \\ 0 & 0 & 0 & \delta_H & 0 \end{pmatrix}$$

Additionally, the characteristic polynomial is as follows:

$$-\lambda\{(-\gamma_V - \mu_V - \lambda)(-\mu_V - \lambda)(-\gamma_H - \lambda)(-\delta_H - \lambda) - \frac{\gamma_V \gamma_H \beta_V \beta_H N^V (N^H - R^H)}{(N^H)^2}\}$$

$$= \lambda^4 + (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)\lambda^3 + \{(\gamma_V + \mu_V)(\mu_V + \gamma_H + \delta_H) + \mu_V(\gamma_H + \delta_H) + \gamma_H \delta_H\}\lambda^2 + \{(\gamma_V + \mu_V)(\mu_V \gamma_H + \mu_V \delta_H + \gamma_H \delta_H) + \mu_V \gamma_H \delta_H\}\lambda + \{(\gamma_V + \mu_V)\mu_V \gamma_H \delta_H - \frac{\gamma_V \gamma_H \beta_V \beta_H N^V (N^H - R^H)}{(N^H)^2}\}$$

$$= a_4 \lambda^4 + a_3 \lambda^3 + a_2 \lambda^2 + a_1 \lambda + a_0, 0 \leq R^H \leq N^H.$$

Then, when using the criteria of Routh Hurwitz, the characteristic polynomial needs to satisfy the following three conditions.

$a_3 a_2 - a_4 a_1$

$= (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)\{(\gamma_V + \mu_V)(\mu_V + \gamma_H + \delta_H) + \mu_V(\gamma_H + \delta_H) + \gamma_H \delta_H\} - \{(\gamma_V + $

$$\mu_V)(\mu_V\gamma_H + \mu_V\delta_H + \gamma_H\delta_H) + \mu_V\gamma_H\delta_H\}$$

$$= (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V)(\mu_V + \gamma_H + \delta_H) + (\mu_V + \gamma_H + \delta_H)\mu_V(\gamma_H + \delta_H) + (\gamma_H + \delta_H)\gamma_H\delta_H \geq 0, \text{ and}$$

$$a_3a_2a_1 - a_4a_1^2 - a_3^2a_0$$

$$= (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)\{(\gamma_V + \mu_V)(\mu_V + \gamma_H + \delta_H) + \mu_V(\gamma_H + \delta_H) + \gamma_H\delta_H\}\{(\gamma_V + \mu_V)(\mu_V\gamma_H + \mu_V\delta_H + \gamma_H\delta_H) + \mu_V\gamma_H\delta_H\}$$

$$- \{(\gamma_V + \mu_V)(\mu_V\gamma_H + \mu_V\delta_H + \gamma_H\delta_H) + \mu_V\gamma_H\delta_H\}\{(\gamma_V + \mu_V)(\mu_V\gamma_H + \mu_V\delta_H + \gamma_H\delta_H) + \mu_V\gamma_H\delta_H\}$$

$$- (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)^2\{(\gamma_V + \mu_V)\mu_V\gamma_H\delta_H - \frac{\gamma_V\gamma_H\beta_V\beta_H N^V(N^H - R^H)}{(N^H)^2}\}$$

$$= \{(\gamma_V + \mu_V)(\mu_V\gamma_H + \mu_V\delta_H + \gamma_H\delta_H) + \mu_V\gamma_H\delta_H\}\{(\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V)(\mu_V + \gamma_H + \delta_H) + (\mu_V + \gamma_H + \delta_H)\mu_V(\gamma_H + \delta_H) + (\gamma_H + \delta_H)\gamma_H\delta_H\}$$

$$- (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V)\mu_V\gamma_H\delta_H + (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)\frac{\gamma_V\gamma_H\beta_V\beta_H N^V(N^H - R^H)}{(N^H)^2}$$

$$= (\gamma_V + \mu_V)(\mu_V\gamma_H + \mu_V\delta_H + \gamma_H\delta_H)\{(\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V)(\mu_V + \gamma_H + \delta_H) + (\mu_V + \gamma_H + \delta_H)\mu_V(\gamma_H + \delta_H) + (\gamma_H + \delta_H)\gamma_H\delta_H\}$$

$$- (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V)(\gamma_V + \mu_V)\mu_V\gamma_H\delta_H + (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)\frac{\gamma_V\gamma_H\beta_V\beta_H N^V(N^H - R^H)}{(N^H)^2}$$

$$= (\gamma_V + \mu_V)(\mu_V\gamma_H + \mu_V\delta_H)(\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V)(\mu_V + \gamma_H + \delta_H) + (\gamma_V + \mu_V)\gamma_H\delta_H(\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V)(\gamma_H + \delta_H) + (\gamma_V + \mu_V)(\mu_V\gamma_H + \mu_V\delta_H + \gamma_H\delta_H)\{(\mu_V + \gamma_H + \delta_H)\mu_V(\gamma_H + \delta_H) + (\gamma_H + \delta_H)\gamma_H\delta_H\} + (\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)(\gamma_V + \mu_V + \mu_V + \gamma_H + \delta_H)\frac{\gamma_V\gamma_H\beta_V\beta_H N^V(N^H - R^H)}{(N^H)^2} \geq 0, \text{ and}$$

$$a_0 = (\gamma_V + \mu_V)\mu_V\gamma_H\delta_H - \frac{\gamma_V\gamma_H\beta_V\beta_H N^V(N^H - R^H)}{(N^H)^2} \geq 0.$$

That is, $\dfrac{\gamma_V\gamma_H\beta_V\beta_H N^V N^H}{(N^H)^2} - (\gamma_V + \mu_V)\mu_V\gamma_H\delta_H < \dfrac{\gamma_V\gamma_H\beta_V\beta_H N^V R^H}{(N^H)^2}$.

Thus, $\bar{R}^H = N^H - \dfrac{(\gamma_V + \mu_V)\mu_V\delta_H(N^H)^2}{\gamma_V\beta_V\beta_H N^V} < R^H$.

Then, when $R^H > \bar{R}^H$ using the criteria of Routh Hurwitz, all eigenvalues of the characteristic polynomial except one are real negative, where the last eigenvalue is 0. The eigenvector corresponding to $\lambda = 0$, $\{ E^V, I^V, E^H, I^H, R^H \}^T = \{ 0, 0, 0, 0, 1 \}^T$, is the vector tangential to the equilibrium manifold. Thus, each equilibrium point is locally asymp-

totically stable within the invariant set of the dynamical system containing that equilibrium point.

APPENDIX B

SIMULATION RESULTS FOR PARAMETER ESTIMATION IN DIFFERENT SCENARIOS

Table S1: Summary statistics for parameter estimates and M.C.S.E. for 200 simulated samples with various variances; normal error; the bandwidth constant of the kernel smoothing ($\tilde{h} = ch^{5/9}$) is chosen to be 2 in scenario I.

Variance	Param	*True*	DTR Estimate	MCSE	MSE	MADSQ Estimate	MCSE	MSE
	θ_V	0.01	0.012	0.009		0.005	0.003	
0.01	β_V	0.32	0.322	0.010	0.004	0.321	0.003	0.001
	μ_V	0.08	0.082	0.006		0.079	0.002	
	δ_H	0.13	0.142	0.030		0.137	0.006	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.01	0.029	0.036		0.006	0.004	
0.16	β_V	0.32	0.336	0.026	0.089	0.321	0.008	0.001
	μ_V	0.08	0.096	0.029		0.079	0.003	
	δ_H	0.13	0.157	0.057		0.135	0.007	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.01	0.053	0.065		0.007	0.006	
1.0	β_V	0.32	0.356	0.049	0.300	0.323	0.018	0.002
	μ_V	0.08	0.116	0.054		0.080	0.005	
	δ_H	0.13	0.163	0.049		0.134	0.012	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.01	0.049	0.066		0.009	0.009	
2.0	β_V	0.32	0.359	0.049	0.298	0.324	0.028	0.005
	μ_V	0.08	0.114	0.054		0.082	0.008	
	δ_H	0.13	0.168	0.048		0.128	0.017	

Table S2: Summary statistics for parameter estimates and M.C.S.E. for 200 simulated samples with various variances; normal error; the bandwidth constant of the kernel smoothing $(\tilde{h} = ch^{5/9})$ is chosen to be 2 in scenario II.

			DTR			MADSQ		
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.16	0.203	0.079		0.146	0.054	
0.01	β_V	0.54	0.571	0.074	0.093	0.582	0.060	0.025
	μ_V	0.23	0.270	0.074		0.212	0.047	
	δ_H	0.19	0.172	0.068		0.293	0.049	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.16	0.211	0.094		0.151	0.060	
0.16	β_V	0.54	0.581	0.083	0.158	0.576	0.062	0.051
	μ_V	0.23	0.279	0.089		0.217	0.054	
	δ_H	0.19	0.172	0.072		0.271	0.064	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.16	0.212	0.096		0.122	0.065	
1.0	β_V	0.54	0.584	0.084	0.185	0.557	0.061	0.090
	μ_V	0.23	0.280	0.090		0.193	0.058	
	δ_H	0.19	0.168	0.070		0.270	0.060	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.16	0.198	0.101		0.111	0.060	
2.0	β_V	0.54	0.583	0.096	0.217	0.552	0.066	0.79
	μ_V	0.23	0.269	0.092		0.184	0.053	
	δ_H	0.19	0.171	0.073		0.264	0.062	

Table S3: Summary statistics for parameter estimates and M.C.S.E. for 200 simulated samples with various variances; normal error; the bandwidth constant of the kernel smoothing ($\tilde{h} = ch^{5/9}$) is chosen to be 2 in scenario III.

			DTR			MADSQ		
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.03	0.029	0.003		0.065	0.048	
0.01	β_V	0.14	0.145	0.011	0.317	0.162	0.013	3.758
	μ_V	0.06	0.062	0.003		0.094	0.043	
	δ_H	0.11	0.122	0.055		0.140	0.059	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.03	0.032	0.012		0.043	0.043	
0.16	β_V	0.14	0.146	0.013	0.613	0.159	0.013	3.474
	μ_V	0.06	0.064	0.010		0.075	0.039	
	δ_H	0.11	0.123	0.064		0.150	0.063	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.03	0.047	0.042		0.031	0.036	
1.0	β_V	0.14	0.154	0.020	2.315	0.155	0.013	2.206
	μ_V	0.06	0.077	0.037		0.065	0.032	
	δ_H	0.11	0.134	0.067		0.148	0.059	
Variance	Param	*True*	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.03	0.046	0.056		0.027	0.032	
2.0	β_V	0.14	0.156	0.080	3.471	0.153	0.013	1.727
	μ_V	0.06	0.078	0.051		0.061	0.029	
	δ_H	0.11	0.140	0.071		0.148	0.057	

Table S4: Summary statistics for parameter estimates and M.C.S.E. for 200 simulated samples with various variances; normal error; the bandwidth constant of the kernel smoothing ($\tilde{h} = ch^{5/9}$) is chosen to be 2 in scenario IV.

Variance	Param	True	DTR Estimate	MCSE	MSE	MADSQ Estimate	MCSE	MSE
	θ_V	0.21	0.222	0.049		0.046	0.051	
0.01	β_V	0.24	0.236	0.030	0.175	0.148	0.020	3.36
	μ_V	0.23	0.238	0.048		0.076	0.046	
	δ_H	0.22	0.198	0.038		0.126	0.053	
Variance	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.21	0.219	0.075		0.072	0.077	
0.16	β_V	0.24	0.234	0.047	1.042	0.162	0.032	6.723
	μ_V	0.23	0.235	0.073		0.100	0.071	
	δ_H	0.22	0.195	0.061		0.148	0.063	
Variance	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.21	0.214	0.077		0.065	0.066	
1.0	β_V	0.24	0.233	0.049	1.635	0.171	0.106	30.959
	μ_V	0.23	0.231	0.075		0.124	0.166	
	δ_H	0.22	0.195	0.067		0.149	0.066	
Variance	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.21	0.238	0.084		0.070	0.069	
2.0	β_V	0.24	0.244	0.053	2.292	0.177	0.121	30.915
	μ_V	0.23	0.254	0.082		0.129	0.164	
	δ_H	0.22	0.195	0.074		0.147	0.061	

Table S5: Summary statistics for parameter estimates and M.C.S.E. for 200 simulated samples with various variances; normal error; the bandwidth constant of the kernel smoothing ($\tilde{h} = ch^{5/9}$) is chosen to be 2 in scenario V.

Variance	Param	True	DTR Estimate	MCSE	MSE	MADSQ Estimate	MCSE	MSE
	θ_V	0.01	0.011	0.003		0.011	0.001	
0.01	β_V	0.13	0.131	0.008	0.069	0.126	0.005	0.222
	μ_V	0.05	0.046	0.002		0.043	0.002	
	δ_H	0.10	0.105	0.035		0.115	0.001	
Variance	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.01	0.013	0.007		0.011	0.001	
0.16	β_V	0.13	0.132	0.010	0.107	0.128	0.004	0.094
	μ_V	0.05	0.047	0.006		0.044	0.002	
	δ_H	0.10	0.108	0.040		0.115	0.001	
Variance	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.01	0.020	0.027		0.011	0.003	
1.0	β_V	0.13	0.134	0.014	0.417	0.130	0.005	0.035
	μ_V	0.05	0.053	0.024		0.044	0.002	
	δ_H	0.10	0.110	0.036		0.116	0.004	
Variance	Param	True	Estimate	MCSE	MSE	Estimate	MCSE	MSE
	θ_V	0.01	0.018	0.026		0.012	0.004	
2.0	β_V	0.13	0.135	0.013	0.605	0.131	0.005	0.033
	μ_V	0.05	0.052	0.022		0.045	0.003	
	δ_H	0.10	0.114	0.044		0.115	0.005	

Table S6: Summary statistics for parameter estimate and M.C.S.E. of the virus transmission rate from humans to female *Ae. aegypti* (β_V) keeping all others fixed for 200 simulated samples using only the simulated data before the date before the vector control intervention; variance=2.0; normal error; the bandwidth constant of the kernel smoothing $(\tilde{h} = ch^{5/9})$ is chosen to be 2 in all scenarios.

| | | | DTR | | MADSQ | |
Scenario	Param	*True*	Estimate	MCSE	Estimate	MCSE
I	β_V	0.32	0.317	0.017	0.317	0.017
II	β_V	0.54	0.538	0.024	0.538	0.024
III	β_V	0.14	0.156	0.054	0.156	0.054
IV	β_V	0.24	0.223	0.103	0.223	0.104

Table S7: Summary statistics for parameter estimate and M.C.S.E. of the virus transmission rate from humans to female *Ae. aegypti* for 200 simulated samples using only the simulated data before the date before the vector control intervention with various variances; normal error; the bandwidth constant of the kernel smoothing $(\tilde{h} = ch^{5/9})$ is chosen to be 2 in scenario IV.

| | | | DTR | | MADSQ | |
Variance	Param	*True*	Estimate	MCSE	Estimate	MCSE
0.01	β_V	0.24	0.237	0.014	0.237	0.014
0.16	β_V	0.24	0.222	0.054	0.222	0.054
1.0	β_V	0.24	0.204	0.089	0.204	0.089
2.0	β_V	0.24	0.223	0.103	0.223	0.104

APPENDIX C

SIMULATION RESULTS FOR MODEL AND NETWORK COMPARISONS IN DIFFERENT SCENARIOS
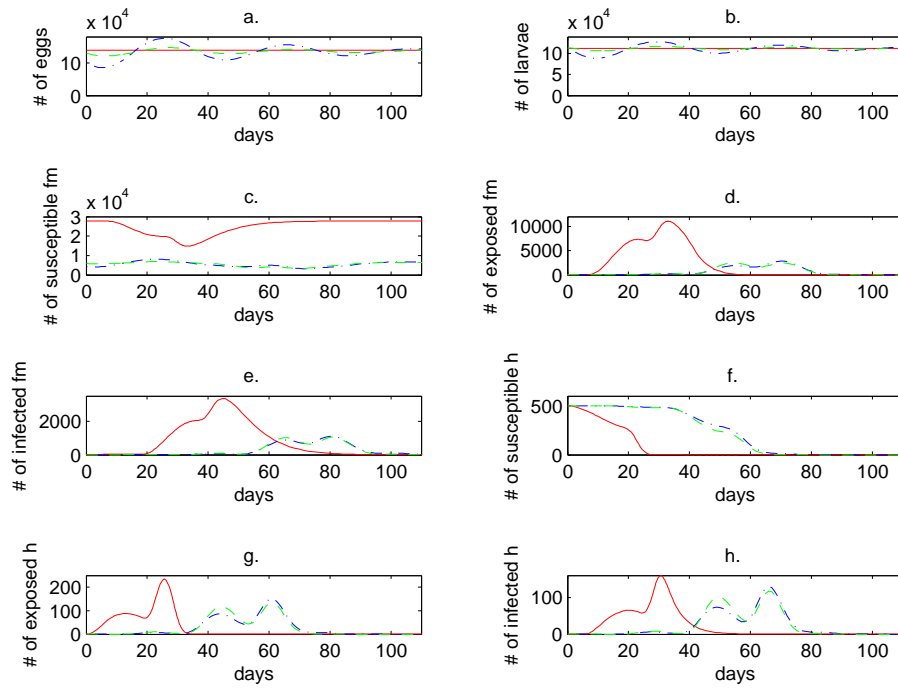
Figure S1: Model comparison between ODE (Red-) and heterogeneous (Blue-.) or homogeneous (Green–) RM network ABM for different state variables when the number of eggs or larvae of *Ae. aegypti* are at the same level
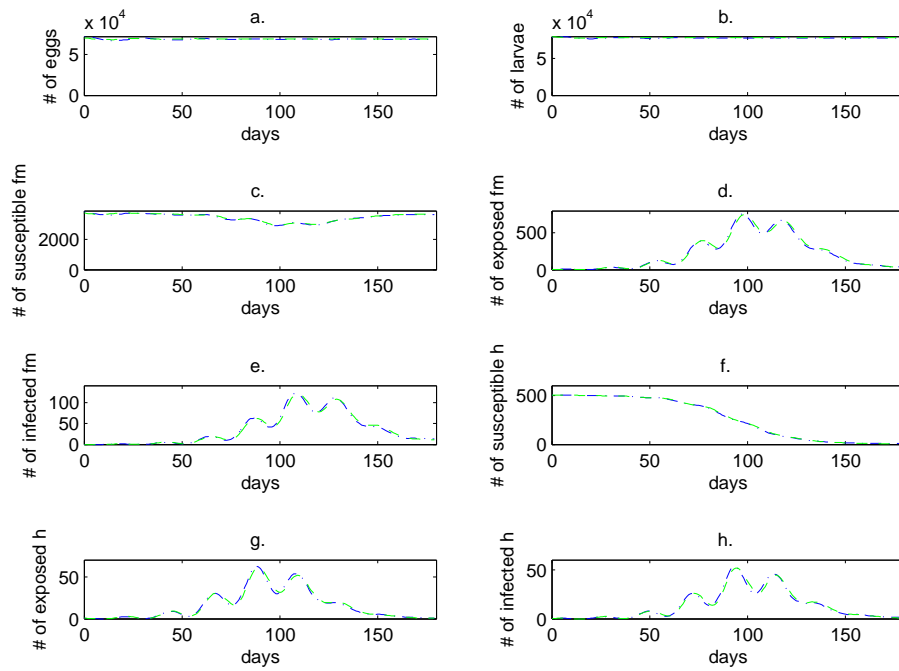
Figure S2: heterogeneous ABM (Blue-.) and homogeneous ABM (Green–) for different state variables in the RM network and OVIP-In structure
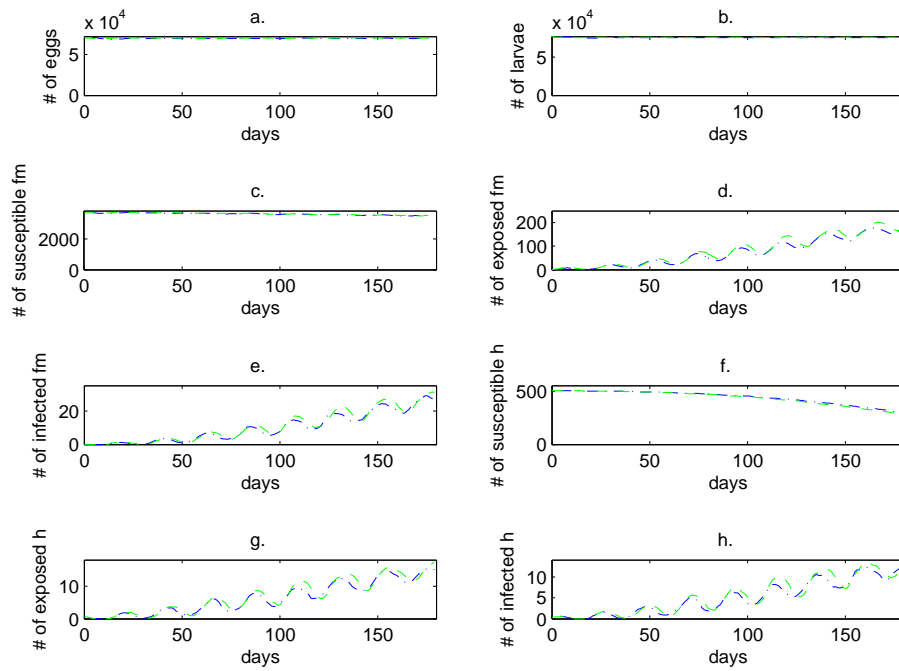
Figure S3: heterogeneous ABM (Blue-.) and homogeneous ABM (Green–) for different state variables in the SW network and OVIP-In structure
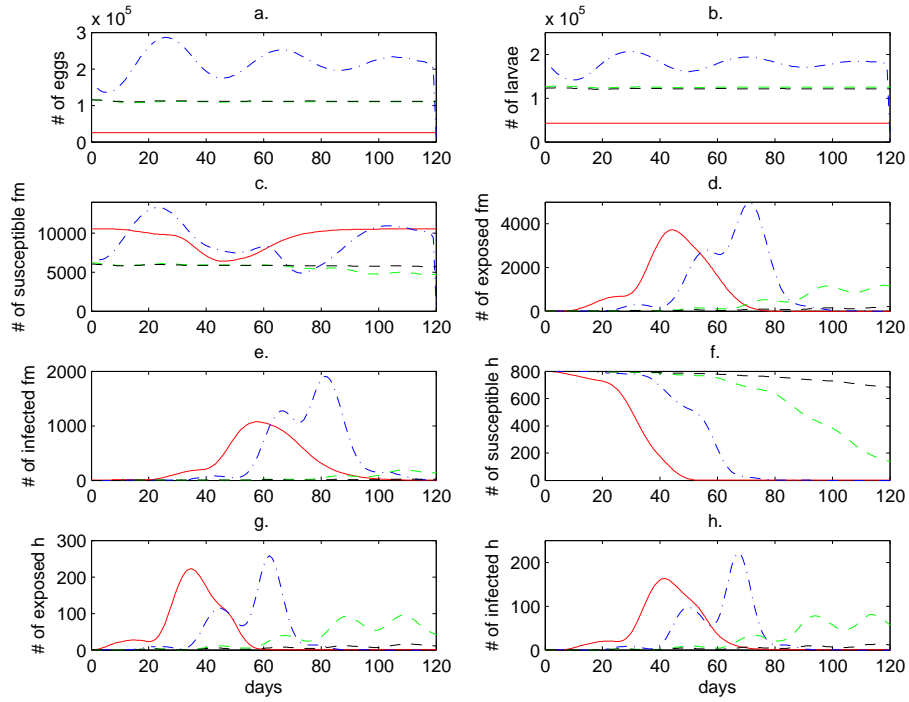
Figure S4: Model or network types comparison with populations of 800 humans for different state variables when the number of total female *Ae. aegypti* are at the same level: ODE curve(Red-) and heterogeneous RM network ABM (Blue-.) in OVIP-Random; heterogeneous RM network (Green–) and SW network (Black–) in OVIP-In
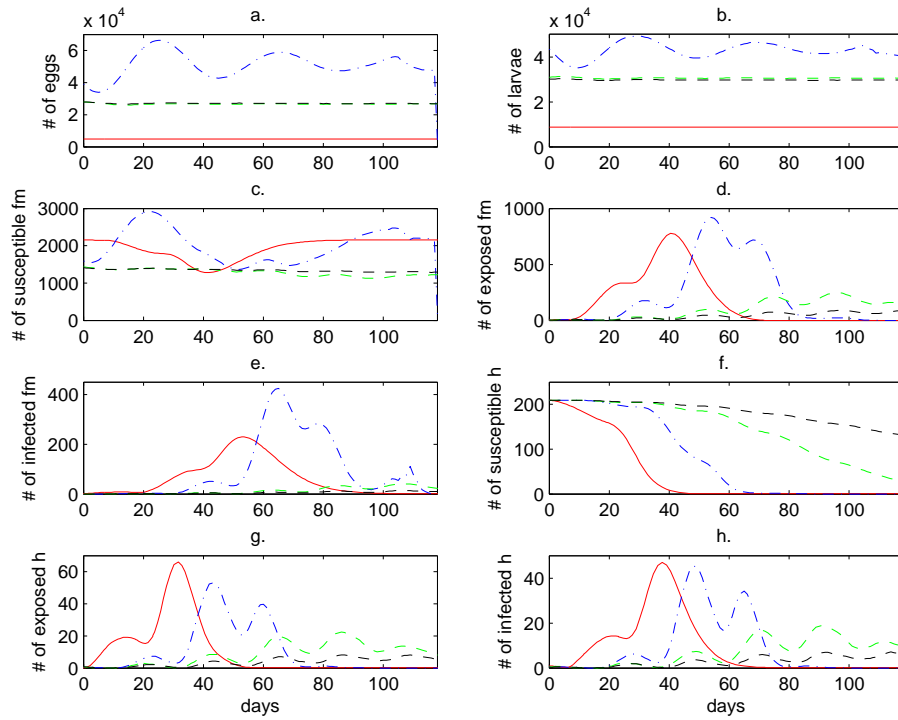
Figure S5: Model or network types comparison with populations of 210 humans for different state variables when the number of total female *Ae. aegypti* are at the same level: ODE curve(Red-) and heterogeneous RM network ABM (Blue-.) in OVIP-Random; heterogeneous RM network (Green–) and SW network (Black–) in OVIP-In

APPENDIX D

MATLAB PROGRAM FOR PARAMETER ESTIMATION

```matlab
%The following settings are for scenario I.
%Scenario II - IV follow the same logic with different settings.

%error term
load('N_124_sim_1_param_4_sd_0.4_error500.mat')
e = Nzsample500;

% ODE System: vSEIR model

ode = @(t,z,p) [ p(1)*(z(1)+z(2)+z(3)) - (p(2)*z(1)*z(6))/2828 - p(3)*z(1);
                (p(2)*z(1)*z(6))/2828 - 0.077*z(2) - p(3)*z(2);
                0.077*z(2) - p(3)*z(3);
                -(p(2)*z(4)*z(3))/2828;
                0.25*(p(2)*z(4)*z(3))/2828 - 0.25*z(5);
                z(5) - p(4)*z(6)];

% Initial Conditions
z0 = [22422; 0; 0; 2827; 0.25; 0];

% Parameter Boundary
lb = [0; 0.1; 0; 0.08];    % Lower bound
ub = [0.4; 1; 1; 0.33];    % Upper bound

% True Parameter Values
p = [0.0101; 0.3176; 0.0815; 0.1304]; %(1/0.077+1/0.25)

% Generate Fitting Data
t   = 0:1:123;                    %measurement times
odeInt = @(t,z) ode(t,z,p);       %ODE function for ODE15s
options = odeset('RelTol',1e-4,'AbsTol',[1e-4 1e-4 1e-4 1e-4 1e-4 1e-4]);
%[~,z] = ode45(odeInt,t,z0,options);  %Can try ode45 first to solve ODEs
[~,z] = ode15s(odeInt,t,z0,options);  %Solve ODEs

% Starting Guess
p0 = [0.011; 0.35; 0.089; 0.143];

%Generate derivatives using symbolic toolbox
%DTR method
%[dfdz,dfdp] = symDynJac(ode,size(z0,1),size(lb,1));

%dopts = optidynset('stateIndex',5,'integrator',...
%'ode15s','dfdp',dfdp,'dfdz',dfdz)
%opt = optiset('solver','mkltrnls','display',...
%'iter','dynamicOpts',dopts);

%MADSQ method. If using DTR, the part should be commented. Vice versa
dopts = optidynset('integrator','ode15s','sensitivity',...
    'none','stateIndex',5);
nomadopts = nomadset('direction_type','ortho n+1 quad',...
    'vns_search',0);
opt = optiset('solver','nomad','solverOpts',nomadopts,...
    'dynamicOpts',dopts);
```

```matlab
% Solve Simulation

%iteration = 200;
iteration = 1; %for demonstration
thetalist = zeros(iteration,size(lb,1));
fvallist = zeros(iteration,1);
exitflaglist = zeros(iteration,1);

for i = 1:iteration
  % Create OPTI Object
  Opt=opti('ode',ode,'data',t,z(:,5)+e(i,:)','z0',z0,...
      'bounds',lb,ub,'theta0',p0,'options',opt);
  [theta,fval,exitflag,info] = solve(Opt);
  thetalist(i,:) = theta;
  fvallist(i,:) = fval;
  exitflaglist(i,:) = exitflag;
end
```

Published with MATLAB® 7.14

BIBLIOGRAPHY

M.A. Abramson and C. Audet. Convergence of mesh adaptive direct search to second order stationary points. *SIAM J. Optim.*, 17:606–609, 2006.

M. Alvarez, R. Rodriquez-Roche, L. Bernardo, S. Vázquez, L. Morier, D. Gonzalez, O. Castro, G. Kouri, S.B. Halstead, and M.G. Guzman. A simplex method for function minimization. *Am. J. Trop. Med. Hyg.*, 75(6):1113–7, 2006.

M. Andraud, N. Hens, C. Marais, and P. Beutels. Dynamic epidemiological models for dengue transmission: A systematic review of structural approaches. *PLoS ONE*, 7(11): e49085, 2012.

M.A. Atkinson, Z. Su, N. Alphey, G.P. Coleman, and L.M. Wein. Analyzing the control of mosquito-borne diseases by a dominant lethal genetic system. *PNAS*, 104(22):9540–9545, 2007.

C. Audet and J.E. Dennis JR. Mesh adaptive direct search algorithms for constrained optimization. *SIAM J. Optim.*, 17:188–217, 2006.

C. Audet, V. Béchard, and S. Le Digabel. Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search. *J. Glob. Optim.*, 41:299–318, 2008.

N.T.J. Bailey. *The mathematical theory of infectious diseases and its applications*. Griffin, London, 1975.

Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, New York, 1974.

L.M. Bartley, C.A. Donnelly, and G.P. Garnett. The seasonal pattern of dengue in endemic areas: mathematical models of mechanisms. *Trans. R. Soc. Trop. Med. Hyg.*, 96:387–397, 2002.

ME Beatty, P Beutels, MI Meltzer, DS Shepard, J Hombach, R Hutubessy, D Dessis, L Coudeville, B Dervaux, O Wichmann, HS Margolis, and JN Kuritsky. Health economics of dengue: a systematic literature review and expert panel's assessment. *Am. J. Trop. Med. Hyg.*, 84:473–488, 2011.

P. Bühlmann. Sieve bootstrap for smoothing in nonstationary time series. *Ann. Stat.*, 26(1): 48–63, 1998.

C. Castillo-Chavez, Z. Feng, and W. Huang. *On the computation of Ro and its role in global stability, Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction.* Springer, 2002.

D.L. Chao, S.B. Halstead, M.E. Halloran, and I.M. Longini Jr. Controlling dengue with vaccines in thailand. *PLoS Negl. Trop. Dis.*, 6(10):e1876, 2012.

G. Chowell, H. Nishiura, and L.M.A. Bettencourt. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J. R. Soc. Interface*, 4:155–166, 2007.

D.V. Clark, M.P. Mammen Jr., and A. Nisalak. Economic impact of dengue fever/dengue hemorrhagic fever in thailand at the family and population levels. *Am. J. Trop. Med. Hyg.*, 72(6):786–791, 2005.

A. Conn, K. Scheinberg, and L. Vicente. *Introduction to Derivative-Free Optimization.* MPS/SIAM Book Series on Optimization. SIAM, Philadelphia, PA, 2009.

A.R. Conn and S. Le Digabel. Use of quadratic models with mesh-adaptive direct search for constrained black box optimization. *J. Optim. Method. Softw.*, 28(1):139–158, 2013.

A.R. Conn, N.I.M. Gould, and P.L. Toint. *Trust-region Methods.* Society for Industrial and Applied Mathematics, 2000.

D.A.T. Cummings, I.B. Schwartz, L. Billings, L.B. Shaw, and D.S. Burke. Dynamic effects of antibody-dependent enhancement on the fitness of viruses. *Proc. Natl. Acad. Sci.*, 102: 15259–15264, 2005.

A. Das, Z. Gao, P.P. Menon, J.G. Hardman, and D.G. Bates. A systems engineering approach to validation of a pulmonary physiology simulator for clinical applications. *J. R. Soc. Interface*, 8(54):44–55, 2011.

C. Deng, H. Tao, and Z. Ye. Agent-based modeling to simulate the dengue spread. In *Proc. SPIE 7143, Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Geo-Simulation and Virtual GIS Environments*, 2008.

M. Derouich, A. Boutayeb, and E.H. Twizell. A model of dengue fever. *Biomed. Eng. Online*, 2(4), 2003.

O. Elerian, S. Chib, and N. Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993, 2001.

T.P. Endy, K.B. Anderson, A. Nisalak, I.K. Yoon, S. Green, A.L. Rothman, S.J. Thomas, R.G. Jarman, D.H. Libraty, and R.V. Gibbons. Determinants of inapparent and symptomatic dengue infection in a prospective study of primary school children in kamphaeng phet, thailand. *PLoS Negl. Trop. Dis.*, 5(3):e975, 2011.

D.A. Focks, E. Daniels, D.G. Haile, and J.E. Keesling. A simulation model of the epidemiology of urban dengue fever: Literature analysis, model development, preliminary validation, and samples of simulation results. *Am. J. Trop. Med. Heg.*, 53:489–506, 1995.

G.F. Fussmann, S.P. Ellner, K.W. Shertzer, and N.G.J. Hairston. Crossing the hopf bifurcation in a live predator-prey system. *Science*, 290:1358–1360, 2000.

A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis.* Chapman and Hall, New York, NY, 2004.

A. Golightly and D.J. Wilkinson. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics and Data Analysis*, 52:1674–1693, 2008.

D.J. Gubler. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.*, 11(3):480–496, 1998.

B. Guy, B. Barrere, C. Malinowski, M. Saville, R. Teyssou, and J. Lang. From research to phase iii: Preclinical, industrial and clinical development of the sanofi pasteur tetravalent dengue vaccine. *Vaccine*, 29:7229–41, 2011.

Y.A. Halasa, D.S. Shepard, and W. Zeng. Economic cost of dengue in puerto rico. *Am. J. Trop. Med. Hyg.*, 86(5):745–752, 2012.

S.B. Halstead. Dengue. *The Lancet*, 370:1644–1652, 2007.

O. Horstick, S. Runge-Ranzinger, M.B. Nathan, and A. Kroeger. Dengue vectorcontrol services: How do they work? a systematic literature review and country case studies. *Trans. R. Soc. Trop. Med. Hyg.*, 104:379–86, 2010.

L.F.O. Jacintho, A.F.M. Batista, T.L. Ruas, M.G.B. Marietto, and F.A Silva. An agent-based model for the spread of the dengue fever: a swarm platform simulation approach. In *SpringSim '10 Proceedings of the 2010 Spring Simulation Multiconference Article No. 2.* Society for Computer Simulation International, 2010.

S. Karl, N. Halder, J.K. Kelso, S.A. Ritchie, and G.J. Milne. A spatial simulation model for dengue virus infection in urban areas. *BMC Infectious Diseases*, 14:447, 2014.

G. Kuno. Review of the factors modulating dengue transmission. *Epidemiol. Rev.*, 17(2): 321–335, 1995.

J.L. Kyle and E. Harris. Global spread and persistence of dengue. *Annu. Rev. Microbiol.*, 62:71–92, 2008.

J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright. Convergence properties of the neldermead simplex method in low dimensions. *SIAM J. Optim.*, 9:112–147, 1998.

S. Le Digabel. Algorithm 909: Nomad: Nonlinear optimization with the mads algorithm. *ACM Transacations on Mathematical Software*, 37(4):44:1–44:15, 2011.

B.Y. Lee, S.B. Connor, D.L. Kitchen, K.M. Bacon, M. Shah, S.T. Brown, R.R. Bailey, Y. Laosiritaworn, D.S. Burke, and D.A. Cummings. Economic value of dengue vaccine in thailand. *Am. J. Trop. Med. Hyg.*, 84(5):764–72, 2011.

M. Legros, A.L. Lloyd, Y. Huang, and F. Gould. Density-dependent intraspecific competition in the larval stage of aedes aegypti (diptera: Culicidae): Revisiting the current paradigm. *J. Med. Entomol.*, 46(3):409–419, 2009.

H. Liang and H. Wu. Parameter estimation for differential equation models using a framework of measurement error in regression models. *J. Am. Stat. Assoc.*, 103(844):1570–1583, 2008.

P.M. Luz, T.N. Lima-Camara, R.V. Bruno, M.G.d. Castro, Sorgine M.H.F., R. Lourenco-de Oliveira, and A.A. Peixoto. Potential impact of a presumed increase in the biting activity of dengue-virus-infected aedes aegypti (diptera: Culicidae) females on virus transmission dynamics. *Mem. Inst. Oswaldo Cruz*, 106:755–758, 2011.

C.M. Macal and M.J. North. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4:151–162, 2010.

B. Marx and P. Eilers. Multidimensional penalized signal regression. *Technometrics*, 47: 13–22, 2005.

I.S. Mbalawata and S. Särkkä. Parameter estimation in stochastic differential equations with markov chain monte carlo and nonlinear kalman filtering. *Comp. Stat.*, 28:1195–1223, 2013.

K.I.M. McKinnon. Convergence of the nelder-mead simplex method to a nonstationary point. *SIMA J. Optim.*, 9(1):148–158, 1998.

J.A. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, 7: 308–313, 1965.

N. Nuraini, E. Soewono, and K.A. Sidarto. Mathematical model of dengue disease transmission with severe dhf compartment. *Bull. Malays. Math. Sci. Soc.*, 30:143–157, 2007.

H. Padmanabha, D. Durham, F. Correa, M. Diuk-Wasser, and A. Galvani. The interactive roles of aedes aegypti super-production and human density in dengue transmission. *PLoS Negl. Trop. Dis.*, 6(8):e1799, 2012.

R. Pearl. The growth of populations. *Quarterly Review of Biology*, 2:532–548, 1927.

L. Perez and S. Dragicevic. An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, 8(50), 2009.

H. Rahmandad and J. Sterman. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, 54(5):998–1014, 2008.

J. O. Ramsay, G. Hooker, D. Campbell, and J Cao. Parameter estimation for differential equations: a generalized smoothing approach. *J. Roy. Statist. Soc. Ser. B (Statistical Methodology)*, 69:741–796, 2007.

R.C. Reiner Jr., S.T. Stoddard, B.M. Forshey, A.A. King, A.M. Ellis, A.L. Lloyd, K.C. Long, C. Rocha, S. Vilcarromero, H. Astete, I. Bazan, A. Lenhart, G.M. Vazquez-Prokopec, V.A. Paz-Soldan, P.J. McCall, U. Kitron, J.P. Elder, E.S. Halsey, A.C. Morrison, T.J. Kochel, and T.W. Scott. Time-varying, serotype-specific force of infection of dengue virus. *Proc. Natl. Acad. Sci.*, 111:2694–2702, 2014.

S.A. Ritchie, A.T. Pyke, S. Hall-Mendelin, A. Day, C.N. Mores, R.C. Christofferson, D.J. Gubler, S.N. Bennett, and A.F. van den Hurk. An explosive epidemic of denv-3 in cairns, australia. *PLoS One*, 8(7):e68137, 2013.

S. Sankaran, C. Audet, and A.L. Marsden. A method for stochastic constrained optimizationusing derivative-free surrogate pattern search and collocation. *J. Comput. Phys.*, 229: 4664–82, 2010.

M. Sciprom, P. Barbazan, and I.M. Tang. Destabilizing effect of the host immune status on the sequential transmission dynamic of the dengue virus infection. *Math. Comput. Modelling*, 45:1053–1066, 2007.

T.W. Scott, G.G. Clark, L.H. Lorenz, P.H. Amerasinghe, P. Reiter, and J.D. Edman. Detection of multiple blood feeding in aedes aegypti (diptera: Culicidae) during a single gonotrophic cycle using a histologic technique. *J. Med. Entomol.*, 30(1):94–9, 1993.

H. Singer. Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended kalman filter and ito-taylor expansion. *Journal of Computational and Graphical Statistics*, 11(4):972–995, 2002.

N.T.B. Stone, D.A.T. Cummings, S.A. Ritchie, D.S. Burke, Y. Weng, and S.T. Brown. Calibration and validation of an individual-based dengue transmission model with multiple interventions. *Submitted to PLoS Negl. Trop. Dis.*

A.K. Supriatna, E. Soewono, and S.A. van Gils. A two-age-classes dengue transmission model. *Math. Biosci.*, 216:114–121, 2008.

S. Syafruddin and M. S. M. Noorani. Seir model for transmission of dengue fever in selangor malaysia. *international Journal of Modern Physics: Conference Series*, 9:380–89, 2012.

J. Timmer. Parameter estimation in nonlinear stochastic differential equations. *Chaos, Solitons and Fractals*, 11(15):2571–2578, 2000.

V.J. Torczon. On the convergence of pattern search algorithms. *SIAM J. Optim.*, 7:1–25, 1997.

A. Tsoularis and J. Wallace. Analysis of the logistic growth models. *Mathematical Biosciences*, 179:21–55, 2002.

G.M. Vazquez-Prokopec, U. Kitron, B. Montgomery, P. Horne, and S.A. Ritchie. Quantifying the spatial dimension of dengue virus epidemic spread within a tripical urban environment. *PLoS Negl. Trop. Dis.*, 4(12):e920, 2010.

A. Wächter and L.T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program., Ser. A*, 106:25–57, 2006.

G. Walker. On periodicity in series of related terms. *Proc. R. Soc. A*, 131:518–32, 1931.

D.M. Watts, D.S. Burke, B.A. Harrison, R.E. Whitmire, and A. Nisalak. Effect of temperature on the vector efficiency of aedes aegypti for dengue 2 virus. *Am. J. Trop. Med. Hyg.*, 36(1):143–152, 1987.

C.R. Williams, P.H. Johnson, T.S. Ball, and S.A. Ritchie. Productivity and population density estimates of the dengue vector mosquito aedes aegypti (stegomyia aegypti) in australia. *Medical and Veterinary Entomology*, 27:313–322, 2013.

D.P. Word, D.A.T. Cummings, D.S. Burke, S. Lamsirithaworn, and C.D. Laird. A nonlinear programming approach for estimation of transmission parameters in childhood infectious disease using a continuous time model. *J. R. Soc. Interface*, 9(73):1983–97, 2012.

H. Xue, H. Miao, and H. Wu. Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Ann. Stat.*, 38(4):2351–2387, 2010.

X. Xun, J. Cao, B. Mallick, R. Carroll, and A. Maity. Parameter estimation of partial differential equation models. *J. Am. Statist. Assoc.*, 108(503):1009–20, 2013.

H.L. Yeap, P. Mee, T. Walker, A.R. Weeks, S.L. O'Neill, P. Johnson, S.A. Ritchie, K.M. Richardson, C. Doig, N.M. Endersby, and A.A. Hoffmann. Dynamics of the "popcorn" wolbachia infection in outbred aedes aegypti informs prospects for mosquito vector control. *Genetics*, 187(2):583–595, 2011.

I.K. Yoon, A.L. Rothman, D. Tannitisupawong, A. Srikiatkhachorn, R.G. Jarman, J. Aldstadt, A. Nisalak, M.P. Jr Mammen, S. Thammapalo, S. Green, D.H. Libraty, R.V. Gibbons, A. Getis, T. Endy, J.W. Jones, C.J.M. Koenraadt, A.C. Morrison, T. Fansiri, C. Pimgate, and T.W. Scott. Underrecognized mildly symptomatic viremic dengue virus infections in rural thai schools and villages. *J. Infect Dis.*, 206(3):389–98, 2012.

G.U. Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philos. Trans. R. Soc. Lond. A*, 226:267–98, 1927.