Comparing vector–host and SIR models for dengue transmission

Abhishek Pandey[a,*], Anuj Mubayi[b], Jan Medlock[c]

[a] *Department of Mathematical Sciences, Clemson University, Clemson, South Carolina 29634, USA*
[b] *Department of Mathematics, Northeastern Illinois University, Chicago, Illinois 60625, USA*
[c] *Department of Biomedical Sciences, Oregon State University, Corvallis, Oregon 97331, USA*

Abstract

Various simple mathematical models have been used to investigate dengue transmission. Some of these models explicitly model the mosquito population, while others model the mosquitoes implicitly in the transmission term. We study the impact of modeling assumptions on the dynamics of dengue in Thailand by fitting dengue hemorrhagic fever (DHF) data to simple vector–host and SIR models using Bayesian Markov chain Monte Carlo estimation. The parameter estimates obtained for both models were consistent with previous studies. Most importantly, model selection found that the SIR model was substantially better than the vector–host model for the DHF data from Thailand. Therefore, explicitly incorporating the mosquito population may not be necessary in modeling dengue transmission for some populations.

Keywords: Dengue, mathematical model, parameter estimation, Markov chain Monte Carlo (MCMC)

1. Introduction

Dengue infection is one of the leading causes of illness in the tropics and subtropics, where it inflicts substantial health, economic and social burdens [1]. Humans are infected with dengue viruses by the bite of an infective female mosquito *Aedes aegypti*, the principal vector of dengue. Once a person gets bitten by an infective mosquito, the virus undergoes an incubation period of about 4 to 7 days, after which

*Corresponding author:Abhishek Pandey,
O-110 Martin Hall, Box 340975,Clemson, SC 29634-0975, USA;
Email: abhishe@g.clemson.edu; *Phone:* +1 (864) 633 9240

the person enters the acute phase of infection. The acute phase can be as short as 2 days and as long as 10 days. If other female *A. aegypti* mosquitoes bite the ill person during this acute phase, those mosquitoes may become infected and subsequently begin the transmission cycle anew. Dengue infection is generally characterized by a sudden onset of fever and other nonspecific signs and symptoms, including frontal headache, body aches, nausea and vomiting [2]. Symptoms range from mild fever to high fever with severe headache and joint pain, and even to internal hemorrhaging, circulatory failure and death. Cases are classified, in order of increasing severity as dengue fever, dengue hemorrahagic fever (DHF) and dengue shock syndrome [3]. Dengue has been recognized in over 100 countries and an estimated 50–100 million cases of dengue fever and several hundred thousand DHF cases occur yearly, depending on epidemic activity [4]. Particularly, in Thailand, dengue disease incidence has increased from 9 per 100 000 in 1958 to 189 per 100 000 in 1998, with the largest reported incidence of 325 per 100 000 in 1987, making dengue a severe public health problem in Thailand [5].

Several mathematical models have been proposed to investigate dengue epidemiology, some of which explicitly model the mosquito population [e.g. 3, 6, 7], while others implicitly model it in the transmission term [e.g. 8–10]. Although both kinds of models have been extensively used for dengue, little guidance exists for which type of model should be preferred. In particular, there has been no comparison of how well these models explain observed incidence. In this study, we considered simple dengue models with and without explicitly modeling mosquitoes, fit both models to DHF incidence data, and used model selection to compare the models.

Fitting models to data validates the model as well as provides estimates of unknown model parameters. There are some examples in the literature where dengue models have been fit to data. Chowell et al. [11] estimated the transmissibility of dengue during a 2002 epidemic in the Mexican state of Colima using municipal epidemic data to evaluate the effect of spatial heterogeneity. Ferguson et al. [12] used longitudinal incidence of serious dengue disease from Thailand and estimated the basic reproductive number R_0 to gain insight into the transmission dynamics and epidemiology of dengue. We fit a simple vector–host dengue model as well as an SIR-type dengue model and obtain estimates of unknown parameters like recovery rate, probability of severe form of disease, mosquito mortality rate, etc.

The goal of the present study is to understand the impact of some modeling assumptions on quantifying estimates of epidemiological metrics for dengue. We applied Bayesian Markov chain Monte Carlo (MCMC) estimation on a simple vector–host dengue model as well as an SIR-type dengue model to estimate model parameters using monthly DHF incidence data in Thailand for January 1984 to March 1985.

The Bayesian MCMC techniques that we used in this study have been commonly used to estimate model parameters of infectious diseases [13–16]. We use the posterior distribution of the model parameters obtained from Bayesian MCMC to perform uncertainty and sensitivity analysis of basic reproductive number R_0 and thereafter, use model selection on a set of vector–host and SIR models to find a model which agrees with the data most parsimoniously.

2. Methods

We built two mathematical models of dengue transmission, one in which the mosquitoes are explicitly tracked and another without explicit mosquito populations. We then used Bayesian MCMC to fit DHF data from Thailand to these two models. In this section, we outline the data source, models and methods and refer to more detailed descriptions in the Supplementary Material.

2.1. Data Source

The Thailand Ministry of Public Health have been recording the number of DHF cases since 1972. Cases are diagnosed using criteria established by the World Health Organization. We obtained the monthly incidence of DHF for Thailand from 1983 to 1997 (Figure 1) [17]. We chose one epidemic, from January 1984 to March 1985 (Figure 2), to fit the dengue models: this particular epidemic was chosen as a clear, representative example among this data. More specifically, we used the cumulative monthly number of DHF cases for the period January 1984 to March 1985. Cumulative incidence is generally smoother than the original incidence data and thus easier to fit and it also easily handles delayed reporting on holidays and weekends.

2.2. Vector–host model

The Ross–Macdonald model, originally developed for malaria, is a standard mathematical model for vector-borne pathogens that tracks infections in both humans and mosquitoes [18]. Following this framework, we built a vector–host model for dengue consisting of three human host compartments, susceptible (the number of susceptible humans is H_S), infectious (H_I) and recovered (H_R), and two mosquito compartments, susceptible (V_S) and infectious (V_I). Mosquitoes do not recover from infection. The

model is the system of differential equations

$$\begin{aligned}
\frac{\mathrm{d}H_S}{\mathrm{d}t} &= B_H - mc\beta_H \frac{V_I}{V} H_S - \mu_H H_S, \\
\frac{\mathrm{d}H_I}{\mathrm{d}t} &= mc\beta_H \frac{V_I}{V} H_S - \gamma_H H_I - \mu_H H_I, \\
\frac{\mathrm{d}H_R}{\mathrm{d}t} &= \gamma_H H_I - \mu_H H_R, \\
\frac{\mathrm{d}V_S}{\mathrm{d}t} &= B_V - c\beta_V \frac{H_I}{H} V_S - \mu_V V_S, \\
\frac{\mathrm{d}V_I}{\mathrm{d}t} &= c\beta_V \frac{H_I}{H} V_S - \mu_V V_I,
\end{aligned} \tag{1}$$

where $H = H_S + H_I + H_R$ and $V = V_S + V_I$ are the human and mosquito population sizes, respectively. A susceptible human gets infected with force of infection $mc\beta_H \frac{V_I}{V}$, where m is number of mosquitoes per person, c is mean rate of bites per mosquito and β_H is the mosquito-to-human transmission probability per bite. Infectious people recover at rate γ_H. The force of infection for mosquitoes is $c\beta_V \frac{H_I}{H}$, where β_V is the human-to-mosquito transmission probability. For simplicity, we ignored disease-induced mortality in both humans and mosquitoes, which is small [19]. Because we only fit the model to an epidemic lasting about a year, we assumed the human population was constant size by using the birth rate $B_H = \mu_H H$. We also assumed the mosquito population was constant size ($B_V = \mu_V V$), neglecting seasonal fluctuations for simplicity.

Standard mathematical analysis of the model (Supplementary Material S1) shows that the basic reproductive number, the number of new human infections caused by a single infected human in an otherwise completely susceptible population, is

$$R_0 = \frac{mc^2\beta_H\beta_V}{\mu_V(\mu_H + \gamma_H)}. \tag{2}$$

In addition, there are two equilibrium points, the disease-free equilibrium and the endemic equilibrium. An equilibrium point is asymptotically stable if nearby orbits converge to it as time increases, and it is globally asymptotically stable if all orbits, not just those nearby, converge to the equilibrium [20]. For $R_0 > 1$, the disease-free equilibrium is unstable and the endemic equilibrium is locally asymptotically stable. The disease-free equilibrium is globally asymptotically stable when $R_0 \leq 1$ (and the endemic equilibrium is out of the relevant state space, having H_I and V_I negative, and unstable).

To simplify the parameter estimation, rather than fitting human mortality rate along with the other parameters, we fixed $\mu_H = 1/69$ y^{-1} based on the average

human duration of life in Thailand in 1984 of about 69 years [21]. The remaining unknown parameters are the human recovery rate (γ_H), the mosquito mortality rate (μ_V), the probability of DHF (p), the mosquito biting rate (c), the number of mosquitoes per person (m), the mosquito-to-human transmission probability (β_H) and the human-to-mosquito transmission probability (β_V). The biting rate, c, always appears in the model multiplied with either β_H or β_V. Similarly, m always appears multiplied with β_H. Therefore, only 2 of these 4 parameters can be separately estimated, which we chose to be $\beta_{aH} = mc\beta_H$ and $\beta_{aV} = c\beta_V$. In addition, the initial proportion of humans recovered in the host population ($h_R(0) = H_R(0)/H$) as well as initial proportion of mosquitoes infected in the vector population ($v_I(0) = V_I(0)/V$) are unknown and must be determined. Thus, we estimated a total of 5 unknown parameters and 2 initial conditions for the vector–host model. We used the incidence data for January 1984 and Thailand's population in year 1984 to calculate initial conditions for initial proportion of hosts infected, i.e. $h_I(0) = H_I(0)/H(0)$, where $H_I(0) = 454$ and $H(0) = 46\,806\,000$. Since both the human and mosquito populations are constant, initial proportions of susceptible humans and mosquitoes were calculated using the other initial conditions, i.e. $h_S(0) = 1 - h_I(0) - h_R(0)$ and $v_S(0) = 1 - v_I(0)$.

2.3. SIR model

Dengue transmission has been extensively modeled using SIR-type models, which only explicitly track human infections [e.g. 8–10]. These SIR models are simpler than vector–host models, making analysis and parameter estimation easier. SIR models for dengue have typically been constructed directly [e.g. 8]. Alternately, an SIR model can be derived from a vector–host model by assuming that infection dynamics in the vector are fast compared to those of the host, a quasi-equilibrium approximation [22].

We used a standard SIR model,

$$\begin{aligned} \frac{\mathrm{d}H_S}{\mathrm{d}t} &= B_H - \beta\frac{H_I}{H}H_S - \mu_H H_S, \\ \frac{\mathrm{d}H_I}{\mathrm{d}t} &= \beta\frac{H_I}{H}H_S - \gamma_H H_I - \mu_H H_I, \\ \frac{\mathrm{d}H_R}{\mathrm{d}t} &= \gamma_H H_I - \mu_H H_R, \end{aligned} \tag{3}$$

where $H = H_S + H_I + H_R$ is the human population size. Again, we kept the population size constant by setting the birth rate to $B_H = \mu_H H$. A susceptible person gets infected with force of infection $\beta\frac{H_I}{H}$, where β is the composite human-to-human transmission rate. Comparing the equilibria of the vector–host model and

the SIR model (Supplementary Material S2) provides β in terms of the parameters of the vector–host model:

$$\beta \approx \frac{mc^2\beta_H\beta_V}{\mu_V}. \tag{4}$$

SIR model (3) is a standard mathematical model for *directly* transmitted pathogens like influenza and has been thoroughly analyzed [e.g. 23]. The basic reproductive number is

$$R_0 = \frac{\beta}{\mu_H + \gamma_H}. \tag{5}$$

As with vector–host model (1), there are two equilibrium points, the disease-free equilibrium and the endemic equilibrium: for $R_0 > 1$ the disease-free equilibrium is unstable and the endemic equilibrium is globally stable, while the disease-free equilibrium is globally asymptotically stable for $R_0 \leq 1$ (with the endemic equilibrium having $H_I < 0$ and being unstable).

The unknown parameters are the transmission rate (β), the recovery rate (γ_H) and the probability of DHF (p), along with the initial proportion of humans recovered ($h_R(0) = H_R(0)/H$). As in the vector–host model, we used the fixed value for the human mortality rate $\mu_H = 1/69$ y^{-1} to simplify the parameter estimation. Again, the initial proportion of infected humans is given by $h_I(0) = H_I(0)/H(0)$ with $H_I(0) = 454$ and $H(0) = 46\,808\,000$. Like the vector–host model, the other initial condition is $h_S(0) = 1 - h_I(0) - h_R(0)$.

2.4. Bayesian Markov chain Monte Carlo estimation

To estimate the unknown parameters, we used a Bayesian MCMC technique. Bayesian inference uses prior information of the model parameters from previous studies, which is then combined with new data to generate estimates in the form of a probability distribution for the parameters. More precisely, for parameters θ and data D, with the prior parameter distribution $\Pr(\theta)$ and likelihood function $\Pr(D \mid \theta)$, the posterior parameter distribution $\Pr(\theta \mid D)$ is given by Bayes's Theorem:

$$\Pr(\theta \mid D) = \frac{\Pr(D \mid \theta)\Pr(\theta)}{\Pr(D)} \tag{6}$$

or, alternately,

$$\Pr(\theta \mid D) \propto \Pr(D \mid \theta)\Pr(\theta). \tag{7}$$

Because there are no general closed-form solutions, MCMC or other methods must be used to generate approximate samples from the posterior parameter distribution $\Pr(\theta \mid D)$.

The connection between the data and the parameters is made by the likelihood function $L(\theta) = \Pr(D \mid \theta)$, which is the conditional probability of obtaining the data (D) for the given parameter values (θ). Therefore, $L(\theta)$ needs to be maximized to obtain best-fit parameter set. In our case, the likelihood function is derived from the vector–host and SIR models, the solutions to which provide estimates of the DHF monthly incidence data. We added a compartment to each model to calculate the cumulative number of DHF infections (H_C). We assumed that a fraction p of infections were diagnosed as DHF, with p constant in time. We added differential equations for the H_C compartment,

$$\frac{\mathrm{d}H_C}{\mathrm{d}t} = pmc\beta_H \frac{V_I}{V} H_S, \tag{8}$$

for the vector–host model, and

$$\frac{\mathrm{d}H_C}{\mathrm{d}t} = p\beta \frac{H_I}{H} H_S, \tag{9}$$

for the SIR model, which are precisely the rates of new infections multiplied by p. The "ode15s" function in Matlab was used to numerically solve the vector-host model (1) & (8) and the SIR model (3) & (9). These numerical solutions give the predicted monthly cumulative DHF incidence, $y_i = H_C(t_i)/H$, where $t_i = 0, 30, 60, ...$ days. Using the least-squares error between the cumulative DHF data D_i and the model prediction,

$$E^2 = \sum_{i=1}^{15} \Big(D_i - y_i(\theta) \Big)^2, \tag{10}$$

we assumed the errors were Gaussian, giving the likelihood function

$$L(\theta) = \Pr(D \mid \theta) = \exp\left(-E^2\right). \tag{11}$$

For the prior parameter distributions, we assigned wide uniform distributions, with ranges chosen to represent our general understanding about where the parameter values may lie. In the absence of any information on parameters estimates, we used least-squares fitting to find best-guess estimates of parameters. Estimates of γ_H, μ_V and μ_H from the literature ($\gamma_H = 1/7\ d^{-1}$, $\mu_V = 1/14\ d^{-1}$ and $\mu_H = 1/69\ y^{-1}$) were used and the vector–host model was fitted to the data using least squares in Berkeley Madonna to find initial point estimates $\beta_{aH} = 0.002$, $\beta_{aV} = 1.8$ and $p = 0.04$. We used these initial point estimates to form uniform priors for these parameters such that their point estimates lie inside the range of priors. For the transmission term β of the SIR model, we simply choose a very wide uniform prior.

Where parameters were common to both models, both models used the same prior (Table 1).

To generate the posterior parameter distribution, we used an MCMC method based on the Metropolis algorithm using a Gaussian jumping distribution with an adaptive covariance matrix. For each model, we simulated 4 independent MCMC chains and used the Gelman–Rubin test to determine when the chains had converged to the stationary distribution, i.e. the parameter posterior distribution. The Gelman–Rubin test signals convergence when the variance between independent chains is similar to the variance within the chains. (See Supplementary Material S3 for more details.) Once the Gelman–Rubin test passed, we continued sampling from one of the chains for 10 000 more iterations without updating the covariance matrix, saving every 5th iterate as the posterior parameter distribution.

3. Results

We estimated 7 total parameters for the vector–host model and 4 total parameters for the SIR model by Bayesian MCMC using the cumulative DHF incidence data. Both models with their maximum-likelihood (ML) parameter estimates fit the data well (Figure 3), with the vector–host model fitting slightly better. (More on model fitting and model selection below.)

The estimates of the human recovery rate were similar for both models (Figure 4(d) and Table 1). The average duration of human dengue infection is between 2 and 7 days approximately, with ML estimates of about 2 to 3 days. The initial proportion of humans recovered ($h_R(0)$) was estimated to be small in both models, indicating that the human populations were almost entirely susceptible when the outbreak started.

Estimates of the probability of DHF differed somewhat between models: ML of around 3 DHF cases per 1000 infections from the vector–host model and around 14 DHF cases per 1000 infections from the SIR model. The vector–host model includes several parameters not present in the SIR model. From the vector–host model, the range of average lifespan of mosquitoes ($1/\mu_V$) was found to be approximately 13 to 26 days, with ML estimate of about 15 days. The initial proportion of mosquitoes infected was very small (ML of about 0.5%), so that the outbreak had just started in the mosquitoes as well as the humans.

The transmission rates are not common between the models, but comparison of equilibria of both the models allowed us to compare the composite transmission rate β from the SIR model with $\beta = \beta_{aH}\beta_{aV}/\mu_V$ for the vector–host model (Figure 4(c)). Although, the ML estimates of β from both models are similar, the distribution from

the vector–host model has more weight at higher values of β than the distribution from the SIR model: e.g. the median estimates are 0.4882 and 0.3243 respectively.

The basic reproductive number (R_0), the expected number of secondary cases produced by a single infection in a completely susceptible population, was calculated using equations (2) and (5) for the respective models, for each MCMC parameter sample (Figure 4(i)). For all parameter samples, $R_0 > 1$ as expected since the data show an epidemic, but the R_0 values from the vector–host model (ML: 1.57) are higher than from the SIR model (ML: 1.10).

Because R_0 is an important metric for an infectious disease, we performed uncertainty and sensitivity analysis of R_0 for both models using partial rank correlation coefficients (PRCC). The PRCC measures the independent effect of each input parameter on R_0, assuming the parameters to be independent [24]. The ordering of these PRCCs directly corresponds to the level of statistical influence, the impact that uncertainty in the estimate of a parameter has on the variability of R_0 [25]. We used the "prcc" function of the R library epiR [26].

For both models, all of the parameters were significantly different from 0 (p-value $< 2.5 \times 10^{-135}$). For the vector–host model, all parameters except $h_R(0)$ and $v_I(0)$ were most influential in determining the magnitude of R_0 ($|\text{PRCC}| > 0.5$), while only β and γ_H for the SIR model were most influential on the magnitude of R_0. A positive PRCC value indicates that an increase in that parameter leads to an increase in R_0, while a negative value shows that increasing that parameter decreases R_0. For the parameters that appear explicitly in the R_0 equations (2) and (5), the signs of the PRCCs were as expected. Of the remaining parameters, p and $v_I(0)$ have a negative influence on R_0, while $h_R(0)$ has a positive influence on R_0.

Parameter estimates for both models suggest that the initial proportion of humans recovered and the initial proportion of vectors infectious are very small. As a result, we tried fitting both models by fixing $h_R(0) = 0$ and $v_I(0) = 2h_I(0)$ and estimating the other parameters in order to decrease the complexity of the models (Figure 6). We fit the vector–host model by fixing $h_R(0)$ only, fixing $v_I(0)$ only and fixing both $h_I(0)$ and $v_I(0)$. Similarly we fit the SIR model by fixing $h_R(0)$.

We used the Akaike Information Criterion (AIC) to compare the competing 6 models (Table 2). The AIC is a measure of the relative goodness of fit of a statistical model, balancing fit with number of parameters, finding the simplest model that best approximates the true, but unknown mechanisms generating the data. The SIR model with fixed $h_R(0)$ had the minimum AIC value, implying this model was the best among the models. The difference in AIC between the best model and the others (ΔAIC) gave "considerably less support" for all the vector–host models and "substantial support" for both SIR models [27]. Alternatively, Akaike weights

provide the probability that a model is the best among the set of candidate models. The Akaike weight for the SIR model with fixed $h_R(0)$ gave 62% probability of it being the better model whereas the SIR model where $h_R(0)$ is also estimated was 26% likely to be the better model. There was only a 12% probability that any of the vector–host models was best.

4. Discussion

The fitting of dengue incidence data from Thailand to simple vector–host and SIR model provided estimates of model parameters. The estimates of human recovery rate from both the models suggest a recovery period of 2 to 7 days, which is consistent with the estimates used in previous studies [2, 8]. The estimates of the probability of DHF from the vector–host model and the SIR model are that about 3 and 14 out of 1000 infections develop into DHF, respectively for the two models. Based on the annual number of dengue infections and DHF cases [2], 5 out of 1000 infections develop into DHF.

The ML estimate of the basic reproductive number (R_0) for the SIR model is 30% smaller than the estimate for the vector–host model. This is driven by the recovery rate (γ_H) being estimated as 68% larger in the SIR model. The MLE probability of DHF (p) is 4.9 times larger for the SIR model. The two models—one with high R_0 and low p, the other with low R_0 and high p—both fit the data well. The PRCC result showing a negative influence of p on R_0 confirms the relationship between these two parameters.

Dengue had been causing annual outbreaks in Thailand for some time prior to the 1984 epidemic [5]. Despite this, our estimates of the initial proportion of people immune ($h_R(0)$) from both models are very small. A high birth rate [28] and the reemergence of dengue serotypes 3 and 4 [5] could explain this low immunity. In addition, mosquito seasonality may be important to explain the monthly variation in dengue incidence [6], and keeping the mosquito population constant for simplicity in our model could have contributed towards small estimates of $h_R(0)$.

The vector–host model fits the data slightly better than the SIR model, but the fewer number of parameters results in the SIR model being strongly selected by the AIC. Alternative measures for model selection like the Bayesian Information Criterion and the Deviance Information Criterion more strongly penalize the number of parameters than AIC, so we expect the result of model selection to remain unchanged. This suggests that incorporating mosquito populations explicitly in dengue models may not be necessary to estimate incidence.

We believe that for any vector-borne pathogen, explicitly including vector populations may generally be unnecessary to model prevalence or incidence in human or

other primary host. We expect that models with and without explicit vectors will fit primary-host data about equally well and then the fewer parameters of the model without explicit vectors will result in it being preferred by formal model selection. Other factors like seasonality in mosquito abundance may be crucial to fit some long-term data (e.g. Figure 1), which could result in explicit-vector models fitting the data significantly better than implicit-vector models. In addition, explicit-vector models are necessary when interventions are targeted at the disease vector, e.g. insecticide or genetically modified mosquitoes. When the desired model output is the effectiveness or cost-effectiveness of an intervention that acts on the primary host, our result suggests that implicit-vector models are likely to be sufficient.

The composite transmission parameter (β) for the vector–host model was obtained from the equilibria of the two models and may not be a good approximation for our comparison of the dynamics of the models. This may explain the difference in the estimates of the composite transmission parameter (β) between the two models. This is reinforced by the fact that the SIR model fits the observed data well, but not for the same β values as the vector–host model. Thus, in addition to being preferred by model selection, use of the SIR model is justified when only the equilibrium values are of interest.

We chose to use DHF cases because the data was available monthly, while we are only aware of annually reported DF cases [5]. Moreover, a person infected with DHF is more likely to visit hospital due to the severity of the disease, and so more likely to be diagnosed and reported. Therefore, data on reported DHF cases may be more accurate to actual DHF cases than DF data is to DF cases.

We used a Bayesian MCMC technique for estimation, though other estimation methods have also been used in the literature. In particular, least-squares error fitting is popular [e.g. 29] and the expectation maximization (EM) algorithm has seen some use [30, 31]. We choose Bayesian MCMC approach as it provides a huge amount of modeling flexibility and enables analysis of all the model parameters or functions of parameters. It also has advantage of providing a complete distribution for parameters as posterior distributions instead of point estimates. Moreover, posterior summaries such as mean, medians, maximum likelihoods, maximum, minimum and credible intervals are easy to obtain as well.

In this paper, we fitted dengue incidence data from Thaiand to vector–host and SIR models and obtained estimates of model parameters including average duration of dengue infection in humans, lifespan of mosquitoes and the probability of the severe form of disease. The parameter estimates were consistent with existing published values and PRCC values showed that all the parameters except initial conditions have significant influence on the magnitude of the basic reproduction number R_0. Both the

vector–host model as well as the SIR model fit the incidence data well, however AIC model selection found the SIR model with fixed $h_R(0)$ to be substantially better than the vector–host model, implying that incorporating mosquito population explicitly in a dengue model may not be necessary to explain the incidence data from Thailand.

Acknowledgements

The authors thank Clemson University and Oregon State University for the funding for this study.

References

[1] Center for Disease Control and Prevention, Dengue (Accessed January 14, 2013). URL `http://www.cdc.gov/dengue/`

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

[3] J. Medlock, P. M. Luz, C. J. Struchiner, A. P. Galvani, The impact of transgenic mosquitoes on dengue virulence to humans and mosquitoes, Am. Nat. 174 (4) (2009) 565–577.

[4] M. G. Guzmán, G. Kouri, Dengue: an update, Lancet Infect. Dis. 2 (1) (2002) 33–42.

[5] A. Nisalak, T. P. Endy, S. Nimmannitya, S. Kalayanarooj, R. M. Scott, D. S. Burke, C. H. Hoke, B. L. Innis, D. W. Vaughn, et al., Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999, Am. J. Trop. Med. Hyg. 68 (2) (2003) 191–202.

[6] H. J. Wearing, P. Rohani, Ecological and immunological determinants of dengue epidemics, Proc. Natl. Acad. Sci. U.S.A. 103 (31) (2006) 11802–11807.

[7] L. Esteva, C. Vargas, Influence of vertical and mechanical transmission on the dynamics of dengue disease, Math. Biosci. 167 (1) (2000) 51–64.

[8] D. A. T. Cummings, I. B. Schwartz, L. Billings, L. B. Shaw, D. S. Burke, Dynamic effects of antibody-dependent enhancement on the fitness of viruses, Proc. Natl. Acad. Sci. U.S.A. 102 (42) (2005) 15259–15264.

[9] Y. Nagao, K. Koelle, Decreases in dengue transmission may act to increase the incidence of dengue hemorrhagic fever, Proc. Natl. Acad. Sci. U.S.A. 105 (6) (2008) 2238–2243.

[10] B. Adams, E. C. Holmes, C. Zhang, M. P. Mammen Jr., S. Nimmannitya, S. Kalayanarooj, M. Boots, Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok, Proc. Natl. Acad. Sci. U.S.A. 103 (38) (2006) 14234–14239.

[11] G. Chowell, P. Diaz-Dueñas, J. C. Miller, A. Alcazar-Velazco, J. M. Hyman, P. W. Fenimore, C. Castillo-Chavez, Estimation of the reproduction number of dengue fever from spatial epidemic data, Math. Biosci. 208 (2) (2007) 571–589.

[12] N. M. Ferguson, C. A. Donnelly, R. M. Anderson, Transmission dynamics and epidemiology of dengue: insights from age–stratified sero–prevalence surveys, Philos. Trans. R. Soc. Lond. B Biol. Sci. 354 (1384) (1999) 757–768.

[13] S. Cauchemez, F. Carrat, C. Viboud, A. J. Valleron, P. Y. Boëlle, A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data, Stat. Med. 23 (22) (2004) 3469–3487.

[14] Y. Huang, D. Liu, H. Wu, Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system, Biometrics 62 (2) (2006) 413–423.

[15] J. S. Brownstein, T. R. Holford, D. Fish, Enhancing West Nile virus surveillance, United States, Emerg. Infect. Dis. 10 (6) (2004) 1129–1133.

[16] J. Reiczigel, K. Brugger, F. Rubel, N. Solymosi, Z. Lang, Bayesian analysis of a dynamical model for the spread of the Usutu virus, Stoch. Environ. Res. Risk Assess. 24 (3) (2010) 455–462.

[17] Johns Hopkins Center for Immunization Research, Monthly incidence of Dengue Hemorrhagic Fever, 1983-1997, in 72 provinces of Thailand (Accessed January 14, 2013).
URL `http://web.archive.org/web/20041216074102/http://www.jhsph.edu/cir/dengue.html`

[18] G. Macdonald, The Epidemiology and Control of Malaria, Oxford Univeristy Press, New York, 1957.

[19] M. N. Burattini, M. Chen, A. Chow, F. A. B. Coutinho, K. T. Goh, L. F. Lopez, S. Ma, E. Massad, Modelling the control strategies against dengue in Singapore, Epidemiol. Infect. 136 (3) (2008) 309.

[20] K. T. Alligood, T. D. Sauer, J. A. Yorke, Chaos: An Introduction to Dynamical Systems, Springer, New York, 1996.

[21] World Bank, Life expectancy at birth, total (years) (Accessed January 14, 2013). URL http://data.worldbank.org/indicator/SP.DYN.LE00.IN/countries/TH-4E-XT?page=5&display=default

[22] M. J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, Princeton, 2008.

[23] H. W. Hethcote, The mathematics of infectious disease, SIAM Rev. 42 (4) (2000) 599–653.

[24] S. M. Blower, H. Dowlatabadi, Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example, Int. Stat. Rev. 62 (2) (1994) 229–243.

[25] M. A. Sanchez, S. M. Blower, Uncertainty and sensitivity analysis of the basic reproductive rate: tuberculosis as an example, Am. J. Epidemiol. 145 (12) (1997) 1127–1137.

[26] M. Stevenson, epiR: An R package for the analysis of epidemiological data (Accessed March 10, 2013).
URL http://cran.r-project.org/web/packages/epiR/index.html

[27] K. P. Burnham, D. R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Edition, Springer, New York, 2002.

[28] United Nations, World Fertility Data 2012 (Accessed March 25, 2013).
URL http://www.un.org/esa/population/publications/WFD2012/MainFrame.html

[29] A. Mubayi, C. Castillo-Chavez, G. Chowell, C. Kribs-Zaleta, N. Ali Siddiqui, N. Kumar, Transmission dynamics and underreporting of Kala-Azar in the Indian state of Bihar, J. Theor. Biol. 262 (1) (2010) 177–185.

[30] S. Duncan, M. Gyöngy, Using the EM algorithm to estimate the disease parameters for smallpox in 17th Century London, in: Proceedings of the 2006 IEEE International Conference on Control Applications, Omnipress, Madison, Wisconsin, 2006, pp. 3312–3317.

[31] M. Lavielle, A. Samson, A. Karina Fermin, F. Mentré, Maximum likelihood estimation of long-term HIV dynamic models and antiviral response, Biometrics 67 (1) (2011) 250–259.

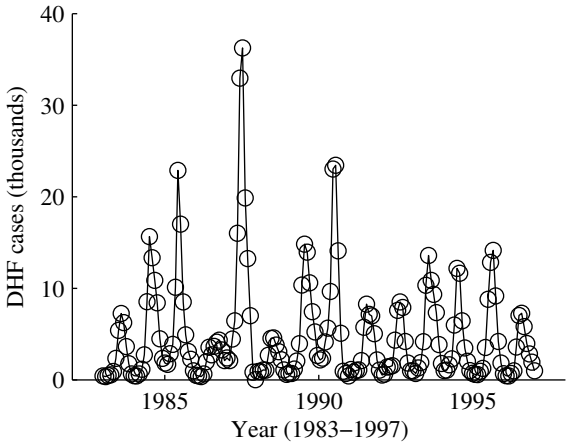

Figure 1: Monthly dengue hemorrhagic fever (DHF) incidence in Thailand from 1983 to 1997.

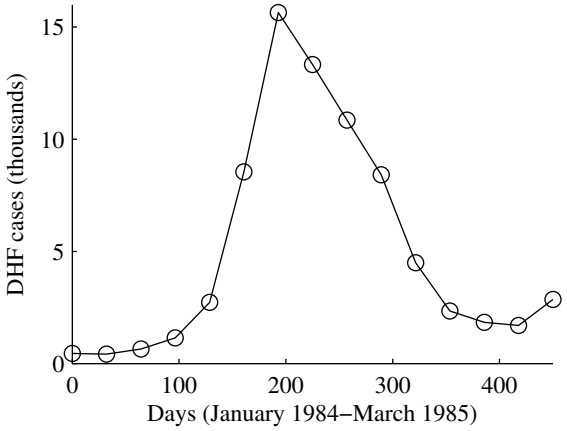

Figure 2: Monthly DHF incidence in Thailand from January 1984 to March 1985.

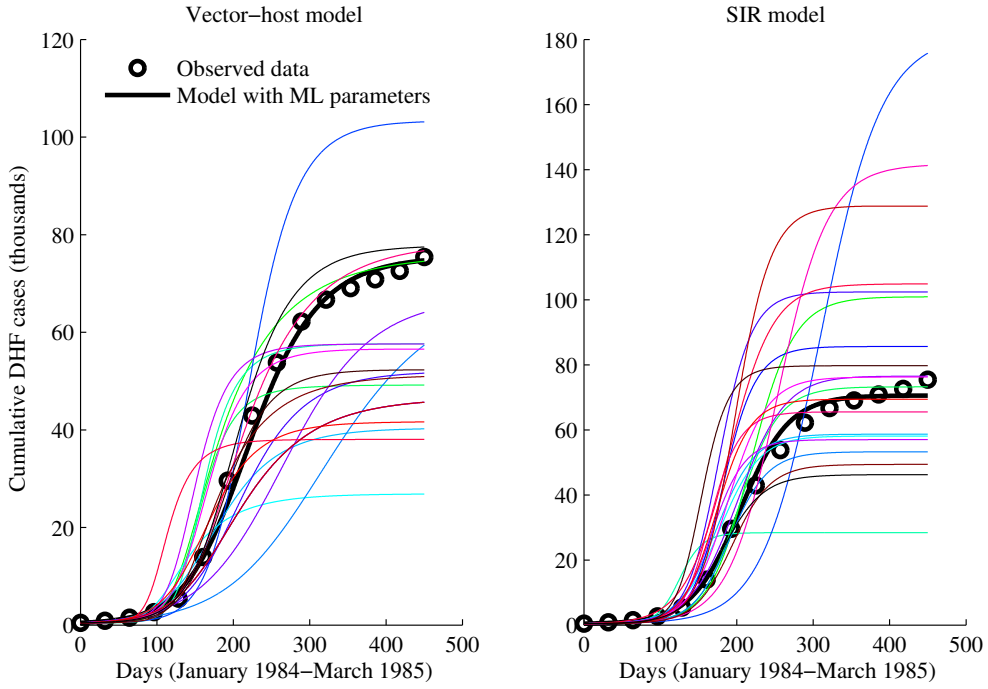

Figure 3: Fits of the vector–host and SIR models. Shown are the cumulative DHF cases from the data (black circles), and from the models with the maximum–likelihood parameter estimates (thick black curves) and 20 samples from the posterior parameter distribution (thin color curves).

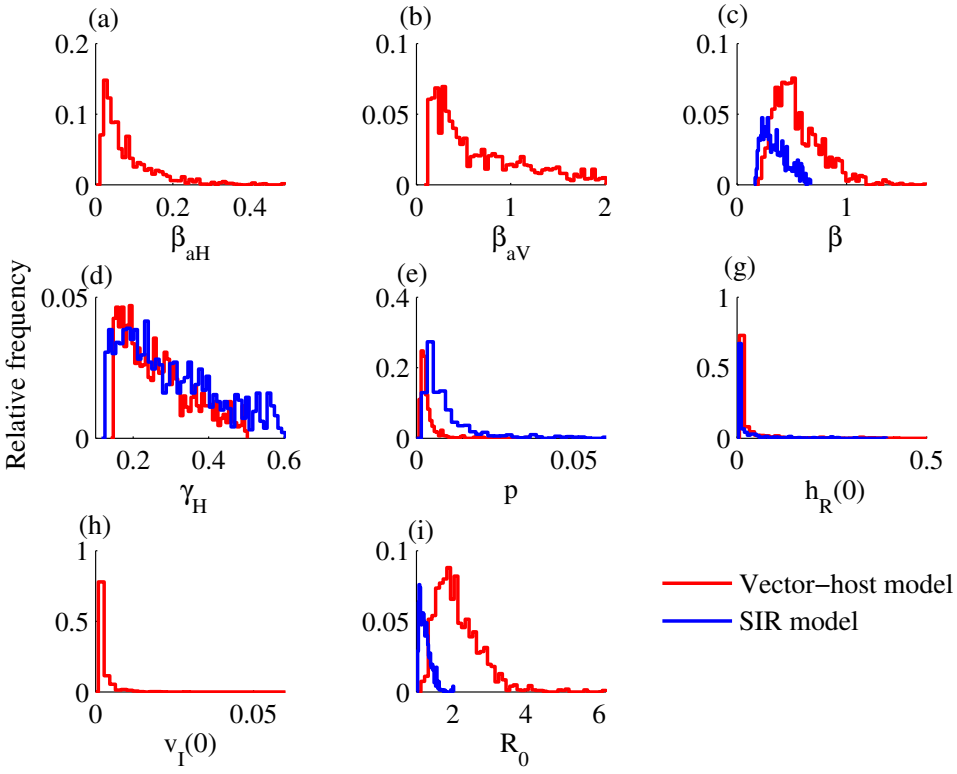

Figure 4: Posterior parameter densities for the vector–host and SIR models. (For the vector–host model, $\beta = \beta_{ah}\beta_{av}/\mu_v$.)

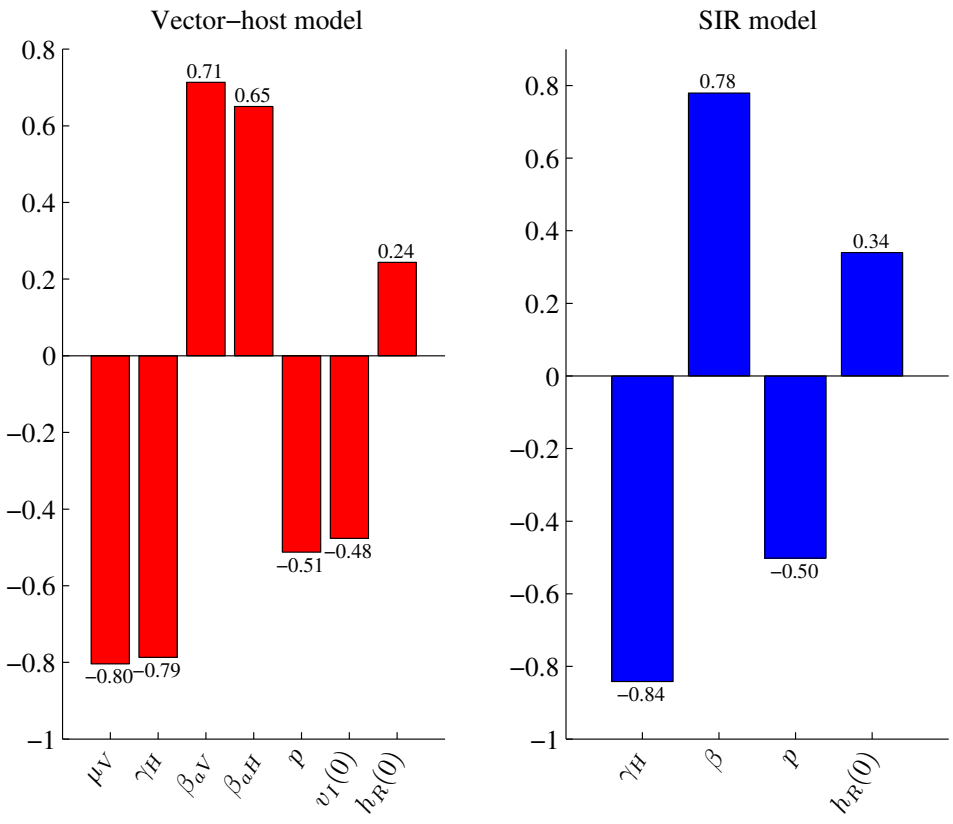

Figure 5: PRCCs for the effect of each parameter on R_0 for the vector–host and SIR models.

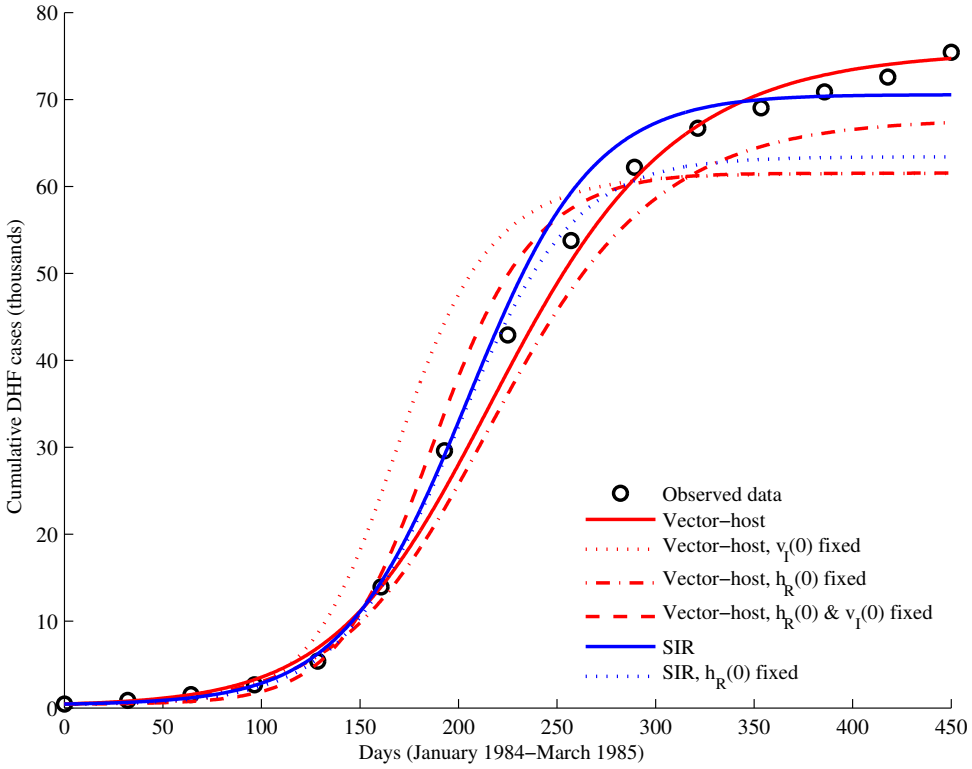

Figure 6: The maximum likelihood fits of models to the data.

Parameter	Prior	Model	Posterior		
			ML	**Median**	**90% CI**
$\beta_{aH}\ (\mathrm{d}^{-1})$ Mosquito-to-human transmission rate	$U(0,1)$	VH	0.0686	0.0521	$(0.0146, 0.2241)$
$\beta_{aV}\ (\mathrm{d}^{-1})$ Human-to-mosquito transmission rate	$U(0.1,2)$	VH	0.4307	0.4867	$(0.1299, 1.6821)$
$\beta\ (\mathrm{d}^{-1})$ Composite transmission rate	$U(0,10)$	VH	0.4881	0.4882	$(0.2782, 0.9364)$
		SIR	0.5718	0.3243	$(0.1931, 0.5805)$
$\gamma_H\ (\mathrm{d}^{-1})$ Human recovery rate	$U(0.1,0.6)$	VH	0.3104	0.2480	$(0.1521, 0.4440)$
		SIR	0.5211	0.2650	$(0.1347, 0.5315)$
p Probability of DHF	$U(0,0.1)$	VH	0.0028	0.0022	$(0.0010, 0.0086)$
		SIR	0.0137	0.0057	$(0.0018, 0.0354)$
$\mu_H\ (\mathrm{y}^{-1})$ Human mortality rate	$1/69$	VH	—	—	—
		SIR	—	—	—
$\mu_V\ (\mathrm{d}^{-1})$ Mosquito mortality rate	$U(0.01,0.1)$	VH	0.0605	0.0531	$(0.0378, 0.0781)$
$h_R(0)$ Initial humans recovered	$U(0,1)$	VH	0.0067	0.0020	$(0.0000, 0.1320)$
		SIR	0.0332	0.0019	$(0.0000, 0.1363)$
$v_I(0)$ Initial mosquitoes infected	$U(0,1)$	VH	0.0009	0.0005	$(0.0000, 0.0056)$
R_0 Basic reproductive number	—	VH	1.5724	1.9733	$(1.3556, 3.2059)$
		SIR	1.0972	1.1989	$(1.0523, 1.5243)$

Table 1: Posterior summary of parameter estimates. For simplicity, μ_h was not estimated. R_0 is not a parameter, but rather a function of the other parameters. $U(a,b)$ is the uniform distribution between a and b. For parameters common to both models, the same prior was used for both models. "ML" is maximum-likelihood estimate; "CI" is credible interval; "VH" is the vector–host model. Parameter units are given in parentheses.

Model	**df**	**Log-likelihood**	**AIC**	**ΔAIC**	**Akaike weight**
Vector–host	7	−0.0779	14.1559	7.7212	0.0130
Vector–host, $v_I(0)$ fixed	6	−0.6847	13.3695	6.9348	0.0193
Vector–host, $h_R(0)$ fixed	6	−0.1173	12.2346	5.7999	0.0341
Vector–host, $h_R(0)$ & $v_I(0)$ fixed	5	−0.5835	11.1669	4.7322	0.0581
SIR	4	−0.1020	8.2039	1.7692	0.2558
SIR, $h_R(0)$ fixed	3	−0.2173	6.4347	0	0.6196

Table 2: Comparison of the vector–host and SIR models with and without fixed initial conditions. "df" is degrees of freedom, i.e. number of parameters.

Supplementary material for *Comparing vector–host and SIR models for dengue transmission*

Abhishek Pandey* Anuj Mubayi† Jan Medlock‡

September 26, 2013

S1 Stability analysis of the vector–host model

For the vector–host model (1), scaling the state variables by their respective population sizes, to the proportions $h_S = H_S/H$, $h_I = H_I/H$, $h_R = H_R/H$, $v_S = V_S/V$ and $v_I = V_I/V$, gives the system of differential equations

$$\begin{aligned}
\frac{\mathrm{d}h_S}{\mathrm{d}t} &= \mu_H - \beta_{aH} v_I h_S - \mu_H h_S, \\
\frac{\mathrm{d}h_I}{\mathrm{d}t} &= \beta_{aH} v_I h_S - \gamma_H h_I - \mu_H h_I, \\
\frac{\mathrm{d}h_R}{\mathrm{d}t} &= \gamma_H h_I - \mu_H h_R, \\
\frac{\mathrm{d}v_S}{\mathrm{d}t} &= \mu_V - \beta_{aV} h_I v_S - \mu_V v_S, \\
\frac{\mathrm{d}v_I}{\mathrm{d}t} &= \beta_{aV} h_I v_S - \mu_V v_I.
\end{aligned} \tag{S1}$$

Using the next-generation method [1], the basic reproductive number is

$$R_0 = \frac{\beta_{aH}\beta_{aV}}{\mu_V(\mu_H + \gamma_H)}. \tag{S2}$$

*Department of Mathematical Sciences, Clemson University, Clemson, South Carolina 29634, USA

†Department of Mathematics, Northeastern Illinois University, Chicago, Illinois 60625, USA

‡Department of Biomedical Sciences, Oregon State University, Corvallis, Oregon 97331, USA

Because $h_S + h_I + h_R = 1$ and $v_S + v_I = 1$, the reduced system

$$\begin{aligned}
\frac{\mathrm{d}h_S}{\mathrm{d}t} &= \mu_H - \beta_{aH} v_I h_S - \mu_H h_S, \\
\frac{\mathrm{d}h_I}{\mathrm{d}t} &= \beta_{aH} v_I h_S - \gamma_H h_I - \mu_H h_I, \\
\frac{\mathrm{d}v_I}{\mathrm{d}t} &= \beta_{aV} h_I v_S - \mu_V v_I,
\end{aligned} \tag{S3}$$

is equivalent to the full system (S1). This system is defined on the domain

$$\Omega = \{(h_S, h_I, v_I) : 0 \leq v_I \leq 1, 0 \leq h_S, 0 \leq h_I, h_S + h_I \leq 1\}. \tag{S4}$$

A simple check shows that the vector field defined by model (S3) on the boundary of Ω does not point to the exterior of Ω, so Ω is positively invariant under the flow induced by system (S3). This guarantees that the model numbers of humans and mosquitoes in the various epidemiological compartments never become negative, which is an obvious biological constraint.

The equilibrium points of system (S3) are

$$E_0 = (1, 0, 0) \qquad \text{and} \qquad E_e = (h_S^*, h_I^*, v_I^*), \tag{S5}$$

where

$$h_S^* = \frac{\delta + M}{\delta + M R_0}, \qquad h_I^* = \frac{R_0 - 1}{\delta + M R_0}, \qquad v_I^* = \frac{\delta(R_0 - 1)}{(\delta + M) R_0}, \tag{S6}$$

with

$$\delta = \frac{\beta_{aV}}{\mu_V} \qquad \text{and} \qquad M = \frac{\mu_H + \gamma_H}{\mu_H}. \tag{S7}$$

E_0 is the disease-free equilibrium and E_e is the endemic equilibrium. For $R_0 < 1$, E_0 is the only equilibrium in Ω but the endemic equilibrium E_e also lies in Ω for $R_0 \geq 1$. The local stability of the equilibrium points is governed by the Jacobian matrix

$$\mathrm{D}F = \begin{bmatrix} -\beta_{aH} v_I - \mu_H & 0 & -\beta_{aH} h_S \\ \beta_{aH} v_I & -(\mu_H + \gamma_H) & \beta_{aH} h_S \\ 0 & \beta_{aV} - \beta_{aV} v_I & -\beta_{aV} h_I - \mu_V \end{bmatrix}. \tag{S8}$$

S1.1 Disease-free equilibrium

The Jacobian matrix (S8) at E_0 is

$$\mathrm{D}F(E_0) = \begin{bmatrix} -\mu_H & 0 & -\beta_{aH} \\ 0 & -(\mu_H + \gamma_H) & \beta_{aH} \\ 0 & \beta_{aV} & -\mu_V \end{bmatrix}, \tag{S9}$$

which has eigenvalues

$$-\mu_H \quad \text{and} \quad \frac{-(\mu_H + \gamma_H + \mu_V) \pm \sqrt{(\mu_H + \gamma_H + \mu_V)^2 - 4\mu_V(\mu_H + \gamma_H)(1 - R_0)}}{2}. \tag{S10}$$

All of the eigenvalues have negative real part for $R_0 < 1$ and so E_0 is locally asymptotically stable for $R_0 < 1$.

To show global stability of E_0, we consider the Lyapunov function on interior of Ω

$$\Lambda = \frac{\beta_{aH}}{\mu_V} v_I + h_I \tag{S11}$$

which has orbital derivative

$$\begin{aligned} \frac{\mathrm{d}\Lambda}{\mathrm{d}t} &= \frac{\beta_{aH}}{\mu_V} \frac{\mathrm{d}v_I}{\mathrm{d}t} + \frac{\mathrm{d}h_I}{\mathrm{d}t} \\ &= -\beta_{aH}(1 - h_S)v_I - (\mu_H + \gamma_H)[1 - R_0(1 - v_I)]h_I. \end{aligned} \tag{S12}$$

For $R_0 \leq 1$, the orbital derivative $\frac{\mathrm{d}\Lambda}{\mathrm{d}t} \leq 0$ in Ω and the subset of Ω where $\frac{\mathrm{d}\Lambda}{\mathrm{d}t} = 0$ is given by

$$(1 - h_S)v_I = 0 \tag{S13}$$

and

$$\begin{aligned} h_I &= 0 \quad \text{if } R_0 < 1, \\ h_I v_I &= 0 \quad \text{if } R_0 = 1. \end{aligned} \tag{S14}$$

Thus $\{E_0\}$ is the only invariant set contained in $\frac{\mathrm{d}\Lambda}{\mathrm{d}t} = 0$. Also, the interior of Ω is bounded. Therefore, E_0 is locally stable and all trajectories starting in Ω approach E_0 as $t \to +\infty$ [2, p. 317, Corollary 1.1]. This establishes the global asymptotic stability of E_0 for $R_0 \leq 1$.

For $R_0 > 1$, the eigenvalue

$$\frac{-(\mu_H + \gamma_H + \mu_V) + \sqrt{(\mu_H + \gamma_H + \mu_V)^2 - 4\mu_V(\mu_H + \gamma_H)(1 - R_0)}}{2} > 0, \tag{S15}$$

so E_0 is unstable.

S1.2 Endemic equilibrium

As R_0 increases through 1, the disease-free equilibrium E_0 becomes unstable and the endemic equilibrium E_e moves from outside to inside Ω. The Jacobian matrix at E_e is

$$\mathrm{D}F(E_e) = \begin{bmatrix} -\mu_H \frac{\delta+MR_0}{\delta+M} & 0 & -\frac{\mu_H M R_0}{\delta}\frac{\delta+M}{\delta+MR_0} \\ \frac{\mu_H M(R_0-1)}{\delta+M} & -\mu_H M & \frac{\mu_H M R_0}{\delta}\frac{\delta+M}{\delta+MR_0} \\ 0 & \frac{\mu_V \delta}{R_0}\frac{\delta+MR_0}{\delta+M} & -\mu_V R_0 \frac{\delta+M}{\delta+MR_0} \end{bmatrix}. \tag{S16}$$

The characteristic polynomial of matrix (S16) is

$$p(\lambda) = \lambda^3 + A\lambda^2 + B\lambda + C, \tag{S17}$$

where

$$\begin{aligned} A &= \mu_H \frac{\delta+MR_0}{\delta+M} + \mu_H M + \mu_V R_0 \frac{\delta+M}{\delta+MR_0} \\ B &= \mu_H^2 M \frac{\delta+MR_0}{\delta+M} + \mu_V \mu_H R_0 + \frac{\mu_H \mu_V M(R_0-1)\delta}{\delta+MR_0} \\ C &= \mu_H^2 \mu_V M(R_0-1). \end{aligned} \tag{S18}$$

For $R_0 > 1$, the coefficients A, B, and C are positive and

$$AB > \mu_H^2 \mu_V M R_0 > C, \tag{S19}$$

so the characteristic polynomial (S17) satisfies the Routh–Hurwithz conditions [3]. Therefore, E_e is locally asymptotically stable.

S2 Comparing equilibria of the vector–host and SIR model

The endemic equilibrium (S6) for the vector–host model has

$$\begin{aligned} h_S^* &= \frac{\delta+M}{\delta+MR_0} = \frac{1+\frac{\mu_H}{\beta_{aH}}R_0}{R_0+\frac{\mu_H}{\beta_{aH}}R_0}, \\ h_I^* &= \frac{R_0-1}{\delta+MR_0} = (R_0-1)\frac{\mu_H}{(\mu_H+\gamma_H)\left(R_0+\frac{\mu_H}{\beta_{aH}}R_0\right)}. \end{aligned} \tag{S20}$$

If

$$\frac{\mu_H}{\beta_{aH}} R_0 \ll 1, \tag{S21}$$

as is true of our ML estimates, then since $R_0 > 1$,

$$\begin{aligned} h_S^* &\approx \frac{1}{R_0}, \\ h_I^* &\approx (R_0 - 1)\frac{\mu_H}{(\mu_H + \gamma_H)R_0} = (R_0 - 1)\frac{\mu_H}{\beta}, \end{aligned} \tag{S22}$$

where

$$\beta = \frac{\beta_{aV}\beta_{aH}}{\mu_V}, \tag{S23}$$

which is exactly the endemic equilibrium of the SIR model (3).

S3 Detailed explanation of the Bayesian MCMC method

Given a mathematical model, the aim is to find the distribution of the unknown parameter values θ for the model that are consistent with the data D. In our case, the data are a time series of monthly cumulative incidence, and $y_i(\theta) = H_C(t_i)/H$ is the time series produced by the model at the same monthly time points t_i as the data. We used the least-squares error function

$$E^2 = \sum_i \Big(D_i - y_i(\theta)\Big)^2 \tag{S24}$$

to calculate the distance between the model output and the data. We assumed that the error function obeys a normal distribution with zero mean and variance 1/2:

$$\Pr\left(E^2\right) \propto \exp\left(-E^2\right). \tag{S25}$$

This is the likelihood function of the parameters given the data,

$$L(\theta) = \Pr(D \mid \theta) \propto \exp\left(-E^2\right). \tag{S26}$$

Once the prior distribution $\Pr(\theta)$ for the unknown parameters is defined, the probability of parameters given the data points, i.e. the posterior distribution, is given by Bayes's Theorem,

$$\Pr(\theta \mid D) = \frac{\Pr(D \mid \theta)\Pr(\theta)}{\Pr(D)}, \tag{S27}$$

where $\Pr(D \mid \theta)$ is the likelihood function (S26) and $\Pr(D)$ is called the evidence, which is the integral of the likelihood over the prior distribution of the parameters:

$$\Pr(D) = \int \Pr(D \mid \theta) \Pr(\theta) \, \mathrm{d}\theta. \tag{S28}$$

It is difficult to compute $\Pr(D)$, but we can instead use

$$\Pr(\theta|D) \propto \Pr(D|\theta) \Pr(\theta) \tag{S29}$$

to find the posterior parameter distribution.

Using Bayes's Theorem (S29), we form a Markov chain which asymptotically converges to the posterior parameter distribution by using the Metropolis Algorithm. The Metropolis Algorithm is an iterative procedure that uses an acceptance–rejection rule to converge to the required distribution [4]. The algorithm is:

1. Start with some initial guess for the parameter values. We randomly chose a starting point θ^0 from the prior distribution $\Pr(\theta)$.

2. For each iteration $n = 1, 2, 3, ...$

   (a) A new proposed set of parameter values is generated by sampling θ^* from the proposal distribution $J\left(\theta^* \mid \theta^{n-1}\right)$. The proposal distribution $J\left(\theta^* \mid \theta^{n-1}\right)$ must be symmetric, i.e. $J\left(\theta^* \mid \theta^{n-1}\right) = J\left(\theta^{n-1} \mid \theta^*\right)$, for this algorithm.

   (b) Using the likelihood function, the ratio

   $$r = \min\left\{\frac{\Pr(\theta^* \mid D) \Pr(\theta^*)}{\Pr(\theta^{n-1} \mid D) \Pr(\theta^{n-1})}, 1\right\} \tag{S30}$$

   is calculated.

   (c) We generate a random uniform number between 0 and 1 and call it α. Then the parameter values for this iteration are

   $$\theta^n = \begin{cases} \theta^* & \text{if } \alpha \ < \ r, \\ \theta^{n-1} & \text{otherwise.} \end{cases} \tag{S31}$$

This algorithm must be run for enough iterations for the parameter values to converge to the posterior distribution. There are several convergence diagnostics which can be employed to detect whether the chain has converged [5]. We used the

Gelman–Rubin test [6] for the convergence diagnostic of our simulations, which is based on multiple independent simulated chains. The variances within each chain are compared to the variances between the chains: large deviation between these two variances indicates non-convergence.

The posterior is insensitive to the choice of proposal density, $J(\theta^* \mid \theta^{n-1})$, but the number of iterations until the chain converges may be heavily affected. It is difficult to choose an efficient proposal distribution, but normal distributions have been found to be useful in many problems [4]. We used a multivariate normal distribution with mean θ^{n-1} and covariance $\lambda^2\Sigma$ as our proposal distribution,

$$J\left(\theta^* \mid \theta^{n-1}\right) \sim N\left(\theta^{n-1}, \lambda^2\Sigma\right). \tag{S32}$$

For the multivariate normal proposal distribution, for optimal convergence, proposals should be accepted at a rate of 0.44 in one dimension and 0.23 in higher dimensions [4]. To achieve this, we used a variant of the Metropolis Algorithm that updates the covariance matrix Σ and the scaling factor λ of the proposal distribution after every 500 iterations. We initially chose the covariance to be the $d \times d$ identity matrix ($\Sigma_0 = I$) and the initial scaling factor to be $\lambda_0 = 2.4/\sqrt{d}$, where d is the number of parameters being estimated. After every 500 iterations, the covariance matrix is updated by

$$\Sigma_k = p\Sigma_{k-1} + (1-p)\Sigma^* \tag{S33}$$

where Σ^* is the covariance of the last 500 parameter values and $p = 0.25$ is the weight given to the old covariance matrix. Similarly, the scaling factor is updated using the Robbins–Monro algorithm [7],

$$\lambda_k = \lambda_{k-1} \exp\left(\frac{\alpha^* - \hat{\alpha}}{k}\right), \tag{S34}$$

where α^* is the acceptance rate for the last 500 iterations and the target acceptance rate is

$$\hat{\alpha} = \begin{cases} 0.44 & \text{for } d = 1, \\ 0.23 & \text{for } d > 1. \end{cases} \tag{S35}$$

We ran the adaptive algorithm in two phases: first, the adaptive phase, which is run until the Gelman–Rubin convergence test passed, and then the fixed phase, where the algorithm is run to sample from the posterior distribution without updating the covariance matrix and the scaling factor. Samples from the only last phase were used in the final inferences.

References

[1] O. Diekmann, J. A. P. Heesterbeek, M. G. Roberts, The construction of next-generation matrices for compartmental epidemic models, J. R. Soc. Interface 7 (47) (2010) 873–885.

[2] J. K. Hale, Ordinary Differential Equations, John Wiley & Sons, New York, 1969.

[3] F. Brauer, C. Castillo-Chávez, Mathematical Models in Population Biology and Epidemiology, Springer, New York, 2001.

[4] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian Data Analysis, 2nd Edition, Chapman & Hall/CRC, Boca Raton, 2004.

[5] M. K. Cowles, B. P. Carlin, Markov chain Monte Carlo convergence diagnostics: a comparative review, J. Am. Statist. Ass. 91 (434) (1996) 883–904.

[6] A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences, Stat. Sci. 7 (4) (1992) 457–472.

[7] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Stat. 22 (3) (1951) 400–407.