HHS Public Access

Author manuscript

Epidemics. Author manuscript; available in PMC 2018 December 01.

Published in final edited form as:

Epidemics. 2018 December ; 25: 89–100. doi:10.1016/j.epidem.2018.05.010.

Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment

Yu-Han Kao[a] and **Marisa C. Eisenberg**[a,b,*]

[a]Department of Epidemiology, University of Michigan, Ann Arbor, MI, United States

[b]Department of Mathematics, University of Michigan, Ann Arbor, MI, United States

Abstract

Mathematical modeling has an extensive history in vector-borne disease epidemiology, and is increasingly used for prediction, intervention design, and understanding mechanisms. Many studies rely on parameter estimation to link models and data, and to tailor predictions and counterfactuals to specific settings. However, few studies have formally evaluated whether vector-borne disease models can properly estimate the parameters of interest given the constraints of a particular dataset. Identifiability analysis allows us to examine whether model parameters can be estimated uniquely—a lack of consideration of such issues can result in misleading or incorrect parameter estimates and model predictions. Here, we evaluate both structural (theoretical) and practical identifiability of a commonly used compartmental model of mosquito-borne disease, using the 2010 dengue epidemic in Taiwan as a case study. We show that while the model is structurally identifiable, it is practically unidentifiable under a range of human and mosquito time series measurement scenarios. In particular, the transmission parameters form a practically identifiable combination and thus cannot be estimated separately, potentially leading to incorrect predictions of the effects of interventions. However, in spite of the unidentifiability of the individual parameters, the basic reproduction number was successfully estimated across the unidentifiable parameter ranges. These identifiability issues can be resolved by directly measuring several additional human and mosquito life-cycle parameters both experimentally and in the field. While we only consider the simplest case for the model, we show that a commonly used model of vector-borne disease is unidentifiable from human and mosquito incidence data, making it difficult or impossible to estimate parameters or assess intervention strategies. This work illustrates the importance of examining identifiability when linking models with data to make predictions and inferences, and particularly highlights the importance of combining laboratory, field, and case data if we are to successfully estimate epidemiological and ecological parameters using models.

Keywords

Dengue; Vector-borne disease; Mathematical modeling; Parameter estimation; Identifiability; Estimability

*Corresponding author at: Department of Mathematics, University of Michigan, Ann Arbor, MI, United States, marisae@umich.edu, (M.C. Eisenberg).

1. Introduction

Arboviral diseases are a global threat of increasing importance. Particularly for diseases propagated by *Aedes* mosquitoes, such as dengue, chikungunya, and Zika (Musso et al., 2015; Benelli and Mehlhorn, 2016), incidences have been increasing at alarming rates worldwide, with over 3.9 billion individuals believed to be at risk for dengue infection alone (Brady et al., 2012; Bhatt et al., 2013; World Health Organization, 2017). These increases are primarily attributed to the habitat expansion of *Aedes spp.* caused by changes in anthropogenic land use and human movement (Wilder-Smith and Gubler, 2008; Guzman et al., 2010; Gubler, 2011; Weaver, 2013; Wilder-Smith, 2012; Powell and Tabachnick, 2013). Given the ecology and life-cycle of *Aedes* mosquitoes, the transmission dynamics of these mosquito-borne diseases are heavily driven by complicated interactions between environmental factors (Hales et al., 2002, 2003; Scott and Morrison, 2003; Chen and Hsieh, 2012; Brady et al., 2014; Kraemer et al., 2015). These factors, combined with human behavior and transmission dynamics, make vector-borne diseases highly complex—presenting both challenges and opportunities for mathematical modeling (World Health Organization, 2011, 2012; Smith et al., 2012). Modeling has increasingly been viewed as a useful tool to quantify these complex transmission systems by integrating various data sources and specifying nonlinear mechanistic relationships and feedbacks. Numerous recent efforts at combating mosquito-borne diseases have directly incorporated the use of mathematical models, such as in planning for Zika and chikungunya response (Moulay et al., 2011, 2012a; Christofferson et al., 2016; Alex Perkins et al., 2016; Kucharski et al., 2016; Ferguson et al., 2016a), and evaluation of potential vaccine candidates (Chao et al., 2012; WHO-VMI, 2012; Aguiar et al., 2016; Ferguson et al., 2016b).

Indeed, mathematical modeling has a long history in vector-borne diseases, beginning with the original development of the Ross-Macdonald or so-called Susceptible-Infectious-Recovered (SIR) model to examine malaria (Kermack and McKendrick, 1927), and expanding to account for an enormous range of factors affecting both human and vector population dynamics (Andraud et al., 2012; Reiner et al., 2013). A wide range of modeling approaches, including ordinary and partial differential equations (ODE and PDE) (Enduri and Jolad, 2014; Aldila et al., 2013) as well as agent/individual-based models have also been applied to these questions (Li and Zou, 2009; Isidoro et al., 2011; Chao et al., 2012; Dommar et al., 2014; Manore et al., 2015). Common goals for many of these modeling efforts have been to make quantitative predictions of disease dynamics and to estimate the underlying mechanistic parameters (Chowell et al., 2007; Khan et al., 2014; Ferguson et al., 2016a; Perkins et al., 2016; Johnson et al., 2015).

To do so often requires using parameter estimation to connect these models with disease data, mainly using incidence or prevalence over time in humans. An important step in this process is examining parameter identifiability, the study of whether a set of parameters can be uniquely estimated and what parameter information may be gleaned from a given model and data set. Unfortunately, under many circumstances, the underlying model parameters are unidentifiable (also denoted non-identifiable), so that many different sets of parameter values produce the same model fit. The unidentifiability (non-identifiability) may be due to

the model and measurement structure (i.e. structural non-identifiability) or the constraints of a specific dataset (i.e. practical unidentifiability). In either case, the data does not provide sufficient information for unique parameter estimation. Incorrect parameter estimates and ignorance of the uncertainty in prediction from an unidentifiable model can result in misleading epidemiological inferences, which could further lead to failures of public health interventions.

There are numerous transmission models in mosquito-borne diseases, which frequently use parameter estimation in fitting these models to data, and broader issues of parameter uncertainty and sensitivity have often been raised and explored (Johnson et al., 2015; Manore et al., 2014; Prosper et al., 2012; Reich et al., 2013; Mendes Luz et al., 2003; Laneri et al., 2010; Pandey et al., 2013; Shutt et al., 2017). However, relatively few efforts have been made to formally examine questions of parameter identifiability in these models (Mendes Luz et al., 2003; Laneri et al., 2010; Bhadra et al., 2011; Moulay et al., 2012b; Pandey et al., 2013; Reiner et al., 2014; Zhu et al., 2015; Tuncer et al., 2016; Shutt et al., 2017). Two studies that directly evaluated identifiability issues include: Denis-Vidal, Verdière, and colleagues assessed the structural (theoretical) identifiability of a chikungunya transmission model assuming all the states in human population and mosquito larva are observable (Moulay et al., 2012b; Zhu et al., 2015); Tuncer et al. (2016) examined both structural and practical identifiability of a within-to-between host model of Rift Valley fever, addressing how the multi-scale nature of such immuno-epidemiological problems affects model identifiability. Building on these results, we examine the identifiability of a simple compartmental model based on the Ross-Macdonald framework with various scenarios of measurement assumption (Newton and Reiter, 1992). This model is commonly used for both theoretical (Coutinho et al., 2005; Coutinhoa et al., 2006; Garba et al., 2008; Dumont et al., 2008) and applied epidemiological studies in a wide range of settings (Burattini et al., 2008; Yang and Ferreira, 2008; Chiroleu and Dumont, 2010; Pinho et al., 2010; Poletti et al., 2011; Sardar et al., 2016), and is often used in an expanded form where temperature or environmental dependence is explicitly included (Bartley et al., 2002; Yang et al., 2009a; Erickson et al., 2010; McLennan-Smith and Mercer, 2014). We consider the structural and practical identifiability of this model in the baseline case without explicit environmental drivers, using dengue incidence data in Kaohsiung, Taiwan as a case study. Additionally, the inclusion of mosquito population data has been considered helpful for parameter estimation in models involving mosquito life cycles (Reiner et al., 2013; Yang et al., 2009a; Morrison et al., 2011; Bowman et al., 2014). However, obtaining mosquito population data is difficult in practice: it requires substantial time and resources which are often limited; spatial and behavioral variability in mosquito populations pose significant logistic challenges as well. Therefore, we also evaluate whether and to what degree that alternative mosquito data available in the field will reduce parameter uncertainty and improve model inference on mosquito control strategies. Finally, we present an example showing the consequences of ignoring unidentifiability in model-based intervention design.

Vector-borne disease modeling is often complex, and has been widely used in forecasting and the design of interventions (Ferguson et al., 2016a; WHO-VMI, 2012; Yakob and Clements, 2013; Patz et al., 1998; Focks and Barrera, 2006; Kearney et al., 2009). Through

our simple model, we hope to draw attention to identifiability issues in vector-borne disease models and their implications in the application of models with more complexity.

2. Methods

In the following sections, we will describe the model development, identifiability analysis, and parameter estimation processes. The flow chart in Fig. 1 summarizes the overall analytical process. The model and analyses were implemented in Python 2.7.10, with code available at https://github.com/epimath/dengue_model.

2.1. Model

Our SEIR-based model is adapted from (Newton and Reiter, 1992; Yang et al., 2009a, 2009b), and shown in Fig. 2. We chose this model mainly because of its simplicity as well as its potential to be used for intervention design and epidemic prediction accounting for environment factors (Yang et al., 2009a, 2009b; Pinho et al., 2010; Poletti et al., 2011; Yakob and Clements, 2013; Erickson et al., 2010; McLennan-Smith and Mercer, 2014; Sardar et al., 2016). The model includes the disease transmission process between the human (h) and mosquito (m) populations. In addition, we specify an aquatic stage of mosquitoes combining larvae and pupae (A). These larvae/pupae then grow into adults (S_m) and leave the aquatic environment. Since dengue virus is transmitted by the female mosquito, we only consider female mosquitoes in the model. The susceptible adult mosquitoes become infected and enter compartment E_m by having blood meals from infectious human beings carrying the dengue virus (I_h). After the extrinsic incubation period (8–12 days) (World Health Organization, 2009; Chan and Johansson, 2012; Rudolph et al., 2014), the infected mosquitoes are capable of transmitting the virus and stay contagious during their lifetime (I_m). Susceptible human individuals (S_h) can be infected (E_h) through bites from the mosquitoes, and then become infectious (I_h) after a 4–10 day intrinsic incubation period (World Health Organization, 2009; Chan and Johansson, 2012; Rudolph et al., 2014). With proper treatment, individuals in the infectious stage can recover from dengue and are considered immune in the model. Note that multiple serotypes are not considered in the model, so potential interactions or antibody-dependent enhancement between serotypes are not included. We assume there is only mosquito-to-human and human-to-mosquito transmission in the model given the relatively low probability of other transmission pathways (World Health Organization, 2009).

2.1.1. Model equations—

In the model, we assume a constant human population ($N = S_h + E_h + I_h + R_h$). We also consider all variables in units of individuals (i.e. humans, mosquitoes, and pupae/larvae).

$$
\begin{aligned}
\frac{\mathrm{dS}_h}{\mathrm{dt}} &= \mu(N - S_h) - \frac{\beta_{mh} S_h I_m}{N} \\
\frac{\mathrm{dE}_h}{\mathrm{dt}} &= \frac{\beta_{mh} S_h I_m}{N} - \alpha E_h - \mu E_h \\
\frac{\mathrm{dI}_h}{\mathrm{dt}} &= \alpha E_h - \eta I_h - \mu I_h \\
\frac{\mathrm{dR}_h}{\mathrm{dt}} &= \eta I_h - \mu R_h \\
\frac{\mathrm{dA}}{\mathrm{dt}} &= \xi(S_m + E_m + I_m)\left(1 - \frac{A}{C}\right) - \pi A - \mu_a A \\
\frac{\mathrm{dS}_m}{\mathrm{dt}} &= \pi A - \frac{\beta_{hm} S_m I_h}{N} - \mu_m S_m \\
\frac{\mathrm{dE}_m}{\mathrm{dt}} &= \frac{\beta_{hm} s_m I_h}{N} - \gamma E_m - \mu_m E_m \\
\frac{\mathrm{dI}_m}{\mathrm{dt}} &= \gamma E_m - \mu_m I_m
\end{aligned}
\tag{1}
$$

It should be noted that β_{mh} and β_{hm} are transmission rates between host and vector populations, which are the products of average bites per mosquito and the probability of successful transmission per infected mosquito bite. C is the maximal carrying capacity of aquatic environment without the additional death term μ_a and maturation rate π. We also include a parameter to account for underreporting in human incidence and prevalence, so that the incidence in the model is measured as $y_h = \kappa_h \alpha E_h$, where κ_h is the reporting fraction. Similarly, for counts and prevalence of mosquitoes, we assume that only a small fraction of the total mosquitoes are counted, assumed to be κ_a and κ_m for aquatic immature and mature mosquitoes, respectively. This yields the (simulated) observed immature mosquitoes to be $y_a = \kappa_a A$ and observed adult mosquitoes to be $y_m = \kappa_m(S_m + E_m + I_m)$. Descriptions of the other parameters are given in Table 1.

2.1.2. Rescaled model—Transmission models such as the one considered here can often be rescaled without changing the observed output. For example, in this model we could rescale the human variables to be larger (thereby also increasing the population size N), but reduce the reporting rate (κ_h) and adjust the value of β_{mh} to yield the same apparent observed number of cases over time from the model. However, because each of these parameters (the reporting rate, transmission parameters, and size of the total population at risk) are all unknown parameters for our model, there is an inherent (structural) unidentifiability of these parameters, so that they cannot all be estimated simultaneously (i.e. for any population size, we can set β_{mh} and the reporting rate to yield the same observed number of cases). Similar issues can be found in the mosquito equations as well.

One way to correct these types of identifiability problems in the model is to rescale the model variables (e.g. S_h, E_h, I_h, S_m, etc.) by model parameters such as the total population size (in many cases this is equivalent to nondimensionalizing the system). In this case, we re-write the human model variables to be in terms of fraction of the population instead of numbers of individuals, e.g. letting the new variable for susceptible humans be: $\tilde{S}_h = S_h/N$

(and similarly for E_h, I_h, and R_h). We also normalize the larvae A by their maximal carrying capacity C (letting $\tilde{A} = A/C$) and the remaining variables (S_m, E_m, and I_m) by both C and π (i.e. letting $\tilde{S}_m = S_m/(C\pi)$) We note that this rescaling of the mosquito variables does not fully nondimensionalize them, but groups parameters into fewer terms (which is useful for identifiability purposes). Rewriting the equations and omitting the ~'s yields:

$$\begin{aligned}
\frac{dS_h}{dt} &= \mu(1 - S_h) - \beta^*_{mh} S_h I_m \qquad (1)\\
\frac{\mathrm{dE}_h}{\mathrm{dt}} &= \beta^*_{mh} S_h I_m - \alpha E_h - \mu E_h\\
\frac{\mathrm{dI}_h}{\mathrm{dt}} &= \alpha E_h - \eta I_h - \mu I_h\\
\frac{\mathrm{dR}_h}{\mathrm{dt}} &= \eta I_h - \mu R_h\\
\frac{\mathrm{dA}}{\mathrm{dt}} &= \xi^*(S_m + E_m + I_m)(1 - A) - \mu^*_a A\\
\frac{\mathrm{dS}_m}{\mathrm{dt}} &= A - \beta_{hm} S_m I_h - \mu_m S_m\\
\frac{\mathrm{dE}_m}{\mathrm{dt}} &= \beta_{hm} S_m I_h - \gamma E_m - \mu_m E_m\\
\frac{\mathrm{dI}_m}{\mathrm{dt}} &= \gamma E_m - \mu_m I_m
\end{aligned}$$

where $\beta^*_{mh} = \beta_{mh} C\pi/N$, $\xi^* = \xi\pi$, and $\mu^*_a = \pi + \mu_a$. Similarly, the reporting rate parameters are now $\kappa^*_h = \kappa_h N, \kappa^*_a = \kappa_a C$, and $\kappa^*_m = \kappa_m C\pi$, so that the observed human cases or mosquito counts are the same as in the original model. However, we note that this means that the reporting rate and population-at-risk can now only be estimated as a combined parameter (as is common for infectious disease models both vector borne and otherwise (Shutt et al., 2017; Eisenberg et al., 2013; Evans et al., 2005)). Rescaling allows us to reduce the number of parameters explicitly included in the model and correct some of the immediately apparent identifiability issues. We will show in Section 2.3 below that this also resolves the overall structural identifiability of the model

For the mosquito population of the rescaled model, we note that interventions involving removal of aquatic phase mosquitoes were conducted once the outbreak began. As a simplified way of representing these dynamics, we used the equilibrium assuming only logistic growth for the aquatic phase mosquitoes to calculate the initial conditions, and then simulated the model assuming an additional death/removal rate μ^*_a (although we note that technically, due to the rescaling, μ^*_a also includes the mosquito maturation rate). This yields initial conditions of 1 for A and $1/\mu_m$ for S_m. These initial conditions also gave a simple way to ensure the aquatic phase of the mosquitoes exhibited some timevarying behavior, since if A was started at the equilibrium using $\mu^*_a \neq 0$, it would remain at equilibrium throughout the simulation (since there is no disease-related death in the mosquito population). We also tested the model when we assumed the initial conditions were at equilibrium with $\mu^*_a \neq 0$ a

(yielding steady state values $A = \frac{\xi^* - \mu_a^* \mu_m}{\xi^*}$ and $S_m = \frac{\xi^* - \mu_a^* \mu_m}{\xi^* \mu_m}$), but this did not change our results (not shown).

2.1.3. Basic reproduction number—The basic reproduction number ($\mathcal{R}_0$) is the total number of secondary cases generated by introducing a single infected individual into a completely susceptible population(vanden Driessche and Watmough, 2002; Heffernan et al., 2005). Mathematically, $\mathcal{R}_0$ is a threshold parameter controlling the stability of the disease-free equilibrium given by an entirely susceptible human and mosquito population. Using the next generation matrix (van den Driessche and Watmough, 2002), we construct $\mathcal{R}_0$ as:

$$\mathcal{R}_0 = \sqrt{\frac{S_m \alpha \beta_{hm} \beta_{mh} \gamma}{(\alpha + \mu)(\eta + \mu)(\gamma + \mu_m)\mu_m}}. \quad (3)$$

2.2. Parameter estimation

2.2.1. Data—Weekly incidence of dengue cases since 1998 is available from the Taiwan National Infectious Disease Statistics System of Taiwan Centers for Disease Control (CDC) (Centers for Disease Control Taiwan, 2018). Confirmed dengue cases are reported from local hospitals and are released every week to the CDC online platform. In the study, we used 2010 dengue incidence data in Kaohsiung, the main city in southern Taiwan. Dengue outbreaks in Taiwan always start from and are often confined to the south because of the favorable environment for *Aedes spp*. Kaohsiung is usually the main epidemic area during outbreaks, and also has annual outbreaks regularly (Chang et al., 2012). The 2010 epidemic curve of dengue in Kaohsiung is very typical with one main peak. Since our model does not handle spatial heterogeneity and multiple strains, we chose to focus only on the 2010 data in Kaohsiung for these analyses.

2.2.2. Parameter estimation—We neglect population birth/death dynamics in the model (μ=0) because the out break only lasts for 32 weeks. We also fix α and γ as 0.14 and 0.1 respectively based on previous studies (World Health Organization, 2009; Chan and Johansson, 2012; Rudolph et al., 2014), and let η be 0.1 since the infection usually lasts for about 10 days (World Health Organization, 2009). We estimated the remaining 6 parameters using weekly dengue incidence in Kaohsiung with least squares (i.e. maximum likelihood assuming normally distributed measurement errors). Nelder-Mead from NumPy in Python 2.7.10 was used for the estimation process.

2.2.3. Simulated data—As discussed in identifiability analysis below, we also simulated noise-free data using the fitted model from previous step. These data were generated by simulating the given variables at either daily or weekly frequency. This allowed us to examine identifiability of the model in a case where the “true” parameters are known (so that errors in estimation can be assessed) and to consider a range of alternative measurement scenarios examining how adding different types of mosquito count data might improve parameter identifiability. We synthesized the following four alternative simulated

data sets corresponding to different surveillance methods available in the field—dengue incidence, ovitrap/house index, BG-trap, and Gravid trap, respectively:

- Scenario 1: human incidence data only, given by $y_h = \kappa_h \alpha E_h$ (integrated to a weekly cumulative incidence)
- Scenario 2: human incidence data (y_h) and daily aquatic (immature) mosquito counts, given by $y_a = \kappa_a A$
- Scenario 3: human incidence data (y_h), aquatic mosquito counts (y_a), and daily adult mosquito counts, given by $y_m = \kappa_m (S_m + E_m + I_m)$
- Scenario 4: human incidence data (y_h),aquatic mosquito counts (y_a), and daily adult mosquito counts broken down by infection status, allowing us to break y_m into $y_{ms} = \kappa_m S_m$ and $y_{mei} = \kappa_m (E_m + I_m)$.

We simulated these data in their most optimistic, best-case form—frequent measurements without noise. However, to examine how noise might affect the parameter estimation, we also generated 300 simulated data sets for Scenario 1 with added measurement error based on the residuals from the parameter estimation with the Kaohsiung data (see Supporting Information).

2.2.4. Estimation with simulated data—For parameter estimation using the simulated data, we fit the model with weighted least squares to account for the different scales for mosquito and human data sets. The weights are the same for each point within each individual dataset (i.e. weighted by the average data value).

2.3. Identifiability analysis

We evaluated the structural and practical identifiability of the parameters, given the model and different possible data sets described above. We will give a brief overview of the identifiability definitions and methods used here. For a more complete review, please refer to (Cobelli and DiStefano, 1980; Audoly et al., 2001; Raue et al., 2009; Miao et al., 2011).

In general there are two types of identifiability: *structural identifiability* (sometimes just called *identifiability*), which examines the best-case scenario of perfectly measured, noise-free data, in order to reveal the inherent, theoretical identifiability derived from the model structure itself; and *practical identifiability* (sometimes called *estimability*), which examines how parameter identifiability fares when real-world data issues such as noise, sampling frequency, and bias are considered (Raue et al., 2009). When a model is unidentifiable, model parameters usually form *identifiable combinations*, which are combinations of parameters that are identifiable even though the individual parameters in the combinations are not.

2.3.1. Structural identifiability analysis—We first examined structural identifiability using two approaches: differential algebra (Ollivier, 1990; Audoly et al., 2001; Pia Saccomani et al., 2001; Meshkat et al., 2012; Eisenberg et al., 2013) and the Fisher information matrix (Rothenberg, 1971; Cobelli and DiStefano, 1980; Cintrón-Arias et al.,

2009; Eisenberg and Hayashi, 2014). A short overview of both methods, formal definitions, and examples are provided in the Supporting Information.

In brief, the differential algebra approach is an analytical method which examines whether it is possible, from the model equations and variables measured, to uniquely determine (estimate) the parameter values. The approach is based only on the model and data structure —it assumes perfect, noise-free data, without consideration of real-world issues of noise, bias, or sampling. This represents an idealized, best-case scenario; however many biological and epidemiological models are structurally unidentifiable, making this a useful first step in examining the parameter information available for a given model and data.

The differential algebra approach provides global results of model structural identifiability and closed forms of the relationships between parameters, but it is usually very computationally expensive. The Fisher information matrix (FIM) can be used as a numerical or analytical approximation to examine structural identifiability for a single point in parameter space (local results), for example, by using very finely sampled simulated data, as discussed in more detail in (Jacquez and Greif, 1985; Eisenberg and Hayashi, 2014). Given that the FIM is often used as a numerical rather than analytical method, there can be limited generalizability across the parameter space. However, it is significantly faster and less computationally intensive than the differential algebra approach.

Here, we test the four simulated data scenarios given above, using the differential algebra approach when possible (using both Mathematica code as well as the freely available packages COMBOS (Meshkat et al., 2014) and Daisy (Bellu et al., 2007)), and the FIM when the differential algebra approach was too computationally intensive to converge to a solution.

2.3.2. Structural and practical identifiability using the profile likelihood—

Another way to assess identifiability is the profile likelihood (Raue et al., 2009). Taking $\mathbf{p} = \{\theta_1, \ldots, \theta_p\}$ as the parameters to be estimated, we fix a parameter (θ_i) across a range of values, which is denoted as $[min(\theta_i), max(\theta_i)]$, and fit the remaining parameters $\{\theta_j | j = 1, \ldots, p, j \neq i\}$ using the likelihood function $\mathcal{L}$ for each value of θ_i in $[min(\theta_i), max(\theta_i)]$. In our case, least squares is used to compute the best-fit values of θ_js, constituting the likelihood profile for the fixed parameter. A minimum in the profile likelihood indicates structural identifiability (at least locally). A parameter is structurally unidentifiable when its likelihood profile is flat and is practically unidentifiable when the curva-ture of its likelihood profile is shallow (Eisenberg and Hayashi, 2014; Raue et al., 2009). However, the degree of shallowness for a profile is a question of degree, so there is often choice of where to set a threshold for practical unidentifiability. In order to decide whether the profile is “flat”, we constructed a 95% upper confidence bound for the profile likelihood given by: $\hat{\sigma}^2 \chi^2_{0.95,p}$ where $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-p}}$ with n denoting the number of observations, p the number of parameters to be estimated, and y and $\hat{y}$ the observations and model trajectory respectively (Raue et al., 2009). Using profile likelihood method, we examine the identifiability of the

model with the four simulated data scenarios as well as the real dengue case data from 2010 in Kaohsiung, Taiwan.

3. Results

3.1. Model fitting and parameter estimation

Using 2010 dengue incidence data in Kaohsiung, the fitted model was able to describe the general trend of the dengue epidemic. The left panel in Fig. 3 shows the dengue incidence data in 2010 and the fitted epidemic curve (y_h). The model captures the overall epidemic size and the long tail at the end (though it overshoots for some of the tail). The fitted parameter values are given in Table 1. Some of the estimated parameters are on the edge of their biologically plausible ranges—for example μ_m, the adult mosquito death rate, corresponds to a mosquito lifespan of approximately three days, which is within the reported range in the literature (Newton and Reiter, 1992; Burattini et al., 2008; Oki et al., 2011) (particularly with ongoing interventions), but short compared to most estimates. However, broadly, the estimated parameter values are difficult to interpret, as the practical unidentifiability of the system (discussed below) means that we can shift the parameters significantly but still achieve the same fit.

As described in the methods, we also simulated both human and mosquito population data which is potentially collectible in the field. The simulated mosquito population data included y_a (aquatic stage), y_m (adult mosquitoes), and y_{ms} (susceptible mosquitoes) and y_{mei} (infected mosquitoes), shown in Fig. 3 (right panel). The fitted model and these simulated data were used for the following identifiability analyses.

3.2. Differential algebra and Fisher information matrix (FIM)

Using the differential algebra approach, we tested the best-case scenario including all the possible data sets from the field, i.e. Scenario 4: dengue incidence, aquatic mosquito counts, infected mosquitoes, and susceptible mosquitoes. With these four types of data together, we proved that the model is structurally identifiable. The detailed proof can be found in the Supporting Information section. However, we were not able to apply the differential algebra method to the remaining three scenarios, due to computational limitations. Therefore, we constructed the FIM to examine the structural identifiability of the model with all scenarios (Scenarios 1–4), using simulated, noise-free dengue incidence and mosquito counts. The FIMs for all the scenarios were full-rank (rank = 6, the number of parameters to be estimated), indicating that the model is locally structurally identifiable at the fitted values in Table 1.

3.3. Profile likelihood of estimated parameters

The parameter profile likelihoods for both the dengue incidence data in 2010, Kaohsiung and the noise-free, simulated incidence data were very similar, with the Scenario 1 profiles shown in Fig. 4 and the Kaohsiung data in Supporting Information Figure S5. Taking β_{mh} in Fig. 4 as an example, the star represents the weighted sum of squared error (SSE) of the original fitted parameter values, and the dots are the SSE after adjusting the β_{mh} value and re-fitting the rest of the parameters. The dashed lines are the thresholds for the approximate

95% confidence bound of the profile likelihood. In principle, the profile likelihood curves of identifiable parameters should cross the thresholds on either side of the minimum (star), and the parameter values where they cross would be the confidence bounds. In this case, all the profiles are flat, meaning the fits are very similar regardless of the changing parameter values, and the confidence bounds are effectively infinite in one or both directions. This result would initially appear at odds with the structural identifiability of the model we showed earlier; however, upon zooming in the profiles, we can see there are minima in each profile (Supporting Information Figure S6). This suggests that although the model is structurally identifiable (consistent with the results from differential algebra and FIM approaches), it is not practically identifiable. To investigate the sources of this practical unidentifiability, we generated scatter plots of each pair of parameters, to evaluate whether any parameters are related to one another and form practically identifiable combinations. We were particularly interested in the pair β_{mh} and β_{hm}—since they form a product in $\mathcal{R}_0$, they could potentially compensate for one another and maintain the same overall magnitude of the epidemic. Indeed, these two parameters do appear to follow an approximate product relationship in their profiles, as illustrated in Fig. 5. In addition, there was a strong linear relationship between ξ and μ_a, which are the parameters controlling the size of aquatic mosquito population. The remaining parameter relationships are shown in Supporting Information.

3.4. Profile likelihood with simulated mosquito data

To evaluate whether including mosquito data collection could enhance model identifiability, we computed profile likelihood of the parameters using simulated mosquito population data sets (Scenarios 2, 3 and 4). A zoomed-in comparison between the β_{mh} profiles of Scenario 1 (only human incidence data), Scenario 2 (adding larva data), Scenario 3 (adding larva and adult mosquito data), and Scenario 4 (adding larva, adult mosquito and infected mosquito data) is shown in Fig. 6. The profiles were improved after adding mosquito information, as the curve slightly tilts up on the right-hand side and becomes higher on the left-hand side. However, the profiles including mosquito population data still do not exceed the 95% confidence threshold within a very wide range of β_{mh}, implying that in practice there is not much obvious improvement on the profile likelihood after including mosquito surveillance data (Supporting Information Figure S4). We note that the small deviations from the profile curve are due to non-convergence of the estimation algorithm for some runs. The profiles for the remaining parameters are similar and are given in (Supporting Information Figure S4). The one exception to the overall trend of practical unidentifiability was that the reporting fraction parameter for the immature mosquitoes (κ_a) was identifiable for all scenarios where mosquito data is measured (this parameter does not appear when only human data is used). The results using simulated noisy data in Scenario 1 (human data only) were very similar to those without noisy data, also showing flat profiles that did not reach the threshold for finite confidence bounds (see Supplementary Figure S3).

3.5. Profile likelihood with fixed parameters

Another way to resolve practical unidentifiability is to decrease the number of parameters to be estimated, which can be done in the real world by having more information about specific parameters, such as using laboratory data to estimate the death rate for mosquito larvae. We

examined this situation by fixing different sets of parameters to their originally fitted values (Table 1) and fitting the remaining parameters using synthesized dengue incidence data (Scenario 1). We demonstrate the results for the β_{mh} profile likelihood in Fig. 7. Given the relationship between β_{hm} and β_{mh}, one might expect fixing β_{hm} could resolve β_{mh}'s identifiability; nevertheless, the profiles indicate that fixing only one of the parameters appearing in $\mathcal{R}_0$ (β_{hm} or μ_m) is not sufficient to make β_{mh} identifiable. Fixing any of other combinations of the parameters not shown in $\mathcal{R}_0$ does not improve β_{mh}'s identifiability either. However, after fixing β_{hm} as well as either μ_m or the pair ξ and μ_a, we obtained profile likelihoods with clear minima, crossing the confidence interval threshold, suggesting with a better idea or prior knowledge about these parameters, we can make β_{mh} identifiable. Unfortunately, as shown in Supporting Information, the whole model does not become identifiable until we fix at least four out of six parameters of interest.

The relatively small number of parameters in this model made it possible to near-exhaustively test a subsets of parameters to determine which ones yielded model identifiability. We started here with β_{mh} given the strong and apparent combination structure between the two transmission parameters (Fig. 5), and then tested fixing increasing subsets of parameters until we found subsets that resulted in finite confidence bounds. However, for larger models with more parameters, a more systematic approach would be needed, such as those presented in (Cintrón-Arias et al., 2009; Eisenberg and Hayashi, 2014; Balsa-Canto et al., 2010; Chis et al., 2011; Brun et al., 2002). A similar idea could also be incorporated in a Bayesian framework by adding sufficiently strong priors to some of the unidentifiable parameters, which could allow successful estimation of the parameters. Indeed, this may be preferable in a real-world setting where parameter priors could be derived from the uncertainty in measuring the parameters through experimental/ecological studies. We note that due to the model unidentifiability, the estimation would thus rely heavily on the priors.

3.6. Basic reproduction number ($\mathcal{R}_0$)

Since $\mathcal{R}_0$ is an important index for understanding disease transmission and predicting future epidemics, a key question is whether we can still estimate $\mathcal{R}_0$ even when the model is practically unidentifiable. As an example exploration of this question, we calculate $\mathcal{R}_0$ using Eq. (3), while profiling parameters β_{mh} and β_{hm}, using Scenario 1 (human incidence data). Fig. 8 demonstrates that $\mathcal{R}_0$ stays stable across the profile of β_{mh} and β_{hm} (the plots of the relationship between $\mathcal{R}_0$ and other parameters are shown in Supporting Information Figure S7. The result indicates that we can often still obtain sensible $\mathcal{R}_0$ estimates from the model with human incidence data, even though we cannot properly estimate the individual parameters.

3.7. Example intervention simulation

We implement a very naive intervention in the model to demonstrate that ignoring unidentifiability can lead to misleading outcomes. We first pick two sets of parameters from the profile in Fig. 4 that generate very similar fits (shown in Fig. 9, left panel). We then remove 10% of the aquatic (immature) mosquito population each day to simulate the population control of mosquito larvae, which is a fairly common countermeasure against dengue. With the same implementation, the responses of the two parameter sets differ

substantially: one epidemic curve only decreases minimally; however, the other simulation decreases significantly and dies out at an early stage of the out-break (Fig. 9, right panel).

4. Discussion

In this study, we explored both structural and practical identifiability of a commonly used SEIR-based model of vector-borne disease. We demonstrated that even when the model is structurally identifiable, it is likely to be difficult or impossible to estimate both human and mosquito parameters from commonly available human incidence data in a single epidemic. In other words, although the likelihood surface of the model has a single optimum, it cannot practically be distinguished from a wide range or curve of neighboring points on the likelihood surface. Moreover, even in cases when human incidence data is combined with the types of mosquito data collected in the field, the practical identifiability of the parameters did not significantly improve. We then showed that more in-depth study of mosquito ecology and behaviors, which can give us direct information about individual parameters, was more efficient in terms of improving model identifiability. Unfortunately, obtaining accurate measurements for any of these parameters individually can be very difficult in practice, as they often vary depending on environmental and ecological factors such as temperature, weather events such as storms, and predation by other species (Kraemer et al., 2015; Wu et al., 2013; Morrison et al., 2008; Chang et al., 2011; Benelli et al., 2016) We would also need additional information on most of the parameters to make the model fully identifiable, which may not always be feasible. Nevertheless, it is still likely possible to measure the relative magnitude of some parameter subsets (e.g. identifiable combinations), providing constraints that can resolve the identifiability issues. For example, one could measure biological factors such as the relative infectivity from mosquitoes to humans and vice versa. This parameter, combined with knowledge of the human and mosquito population sizes (potentially even if only approximately known) could be used to constrain the ratio β_{mh}/β_{hm} and resolve the identifiability of the two β's.

The parameter analyses shown in Fig. 5 and Supplementary Figure S7 give additional guidance on which parameters may resolve the identifiability issues if measured, but more broadly, one could test fixing specific parameter sets using profile likelihoods for a wide range of models. The parameter sets tested might be based on what data is plausible to collect, or one could test parameter sets in a systematic way, e.g. using the identifiable combination structure and/or parameter parameter sensitivites (Cintrón-Arias et al., 2009; Eisenberg and Hayashi, 2014; Balsa-Canto et al., 2010; Chis et al., 2011; Brun et al., 2002). More generally, while we used a maximum likelihood approach here, examining what information was available from human and mosquito surveillance data alone, if parameters were measured using experimental/ecological studies, this information could be included in the model using a Bayesian approach wherein we use the uncertainty in the measured parameters to determine their priors. The fixed-parameter analyses given here can provide some sense of how a very strong prior for the measured parameters would constrain the parameter estimates and reduce uncertainty, but a Bayesian approach with real-world data may provide greater flexibility while still improving identifiability of the system.

In spite of these identifiability problems, the model generates very similar $\mathcal{R}_0$ estimates across a range of profiled parameter values producing the same fit to the data (shown in Fig. 5 and Supplementary Figure S7). This means that estimation using the model may still be useful in characterizing the disease outbreak and spread, calculating vaccination coverage, and assessing the risk of vector-borne disease, even if the individual parameters cannot be determined. $\mathcal{R}_0$ is an important measure that can be used to evaluate potential interventions in public health. For example, we can simulate a model that implements the intervention and compare the $\mathcal{R}_0$ with and without the intervention to evaluate the potential effectiveness (e.g. by examining whether $\mathcal{R}_0$ becomes less than one, or the magnitude of the reduction).

Nevertheless, we cannot solely depend on $\mathcal{R}_0$ since it is possible to obtain very different predicted responses with the same intervention implementation, as shown in (Fig. 9). The two alternative parameter sets shown in Fig. 9 both fit the data equally well and have similar $\mathcal{R}_0$ values (1.30 and 1.33), so that we cannot distinguish which of the predicted intervention responses is more likely. The intervention simulation used here is quite simple, but represents a commonly used control strategy. The example illustrates how a lack of consideration of parameter identifiability can potentially lead to significant errors in evaluating or comparing different intervention strategies.

This model is a simplified interpretation of vectorborne disease transmission, and only assumes one outbreak and a single viral strain. Despite this simple structure, we still cannot properly estimate the parameters from the model. Indeed, similar structural and practical identifiability issues have been noted even for simpler transmission models (Eisenberg et al., 2013; Evans et al., 2005; Cintrón-Arias et al., 2009; Tuncer and Le, 2018). Models with more complicated designs are often more likely to be unidentifiable, underscoring the importance of taking model identifiability into account before making any inferences from the model. Identifiability analysis allows us to understand what a model and data can really tell us, and can help with planning before we invest time and resources into a experimental or field study. Even if unidentifiability is inevitable, as long as we understand the behavior, uncertainty, and the limitations of the model, mathematical models can still be powerful tools to study disease transmission.

In the analyses presented here, we cover a set of basic and often overly optimistic scenarios, simulating data that is noise-free and frequently measured. Given these best-case scenarios, it is unlikely that real world data (which is likely less frequent and noisier) will improve the identifiability of the model. This is illustrated in Supplementary Figure S3 for human surveillance data, and is likely to be similar or worse for mosquito data, which is often noisy, difficult to measure at the daily frequency simulated here, and would also include low prevalences of infection that are difficult to detect. However, more comprehensive research is needed to investigate how issues such as different types of measurement and process noise, missing data, and data resolutions can further complicate parameter estimation. In many cases, these issues will likely further hinder the model parameter estimation and identifiability, but in some cases, more complex dynamics or process noise could potentially improve identifiability, making this a natural next direction for investigation. Nonetheless, this work shows that parameter estimation from incidence data alone is likely to be difficult or impossible, highlighting the importance of integrating parameter information directly

from experimental or field data. Given that such experimentally measured parameters usually vary as a function of environmental variables such as temperature and rainfall (Mordecai et al., 2013, 2017), future work to evaluate how model identifiability changes once this dependence is incorporated into the parameters would be a highly useful next step, particularly as previous studies have shown that uncertainty may vary over different temperature ranges (Johnson et al., 2015).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank Sarah Cherng, Lisa Lau, and Rafael Meza for their helpful comments and discussion of this work. This study was supported by the National Institutes of Health (NIH), National Institute of General Medical Sciences (NIGMS) grant U01GM110712 (MCE and YHK), the International Peace Scholarship from the Philanthropic Educational Organization (YHK), and the Study Abroad Scholarship from the Ministry of Education in Taiwan (YHK).

References

Aguiar M, Stollenwerk N, Halstead SB, 2016 The impact of the newly licensed dengue vaccine in Endemic Countries. PLoS Negl. Trop. Dis 10, e0005179. [PubMed: 28002420]

Aldila D, Götz T, Soewono E, 2013 An optimal control problem arising from a dengue disease transmission model. Math. Biosci 242, 9–16. [PubMed: 23274179]

Alex Perkins T, Siraj AS, Ruktanonchai CW, Kraemer MUG, Tatem AJ, 2016 Model-based projections of Zika virus infections in childbearing women in the Americas. Nat. Microbiol 1, 16126. [PubMed: 27562260]

Andraud M, Hens N, Marais C, Beutels P, 2012 Dynamic epidemiological models for dengue transmission: a systematic review of structural approaches. PLoS ONE 7, e49085. [PubMed: 23139836]

Audoly S, Bellu G, D'Angiò L, Saccomani MP, Cobelli C, 2001 Global identifiability of nonlinear models of biological systems. IEEE Trans. Biomed. Eng 48, 55–65. [PubMed: 11235592]

Balsa-Canto E, Alonso AA, Banga JR, 2010 An iterative identification procedure for dynamic modeling of biochemical networks. BMC Syst. Biol 4, 11. [PubMed: 20163703]

Bartley LM, Donnelly CA, Garnett GP, 2002 The seasonal pattern of dengue in endemic areas: mathematical models of mechanisms. Trans. R. Soc. Trop. Med. Hyg 96, 387–397. [PubMed: 12497975]

Bellu G, Saccomani MP, Audoly S, D'Angiò L, 2007 DAISY: a new software tool to test global identifiability of biological and physiological systems. Comput. Methods Programs Biomed 88, 52–61. [PubMed: 17707944]

Benelli G, Mehlhorn H, 2016 Declining malaria, rising of dengue and Zika virus: insights for mosquito vector control. Parasitol. Res 115, 1747–1754. [PubMed: 26932263]

Benelli G, Jeffries C, Walker T, 2016 Biological control of mosquito vectors: past, present, and future. Insects 7, 52.

Bhadra A, Ionides EL, Laneri K, Pascual M, Bouma M, Dhiman RC, 2011 Malaria in Northwest India: data analysis via partially observed stochastic differential equation models driven by Lévy Noise. J. Am. Stat. Assoc 106, 440–451.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GRW, Simmons CP, Scott TW, Farrar JJ, Hay SI, 2013 The global distribution and burden of dengue. Nature 496, 504–507. [PubMed: 23563266]

Bowman LR, Runge-Ranzinger S, McCall PJ, 2014 Assessing the relationship between vector indices and dengue transmission: a systematic review of the evidence. PLoS Negl. Trop. Dis 8.

Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, Moyes CL, Farlow AW, Scott TW, Hay SI, 2012 Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl. Trop. Dis 6, e1760. [PubMed: 22880140]

Brady OJ, Golding N, Pigott DM, Kraemer MUG, Messina JP, Reiner RC, Jr, Scott TW, Smith DL, Gething PW, Hay SI, 2014 Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. Parasit. Vectors 7, 338. [PubMed: 25052008]

Brun R, Kühni M, Siegrist H, Gujer W, Reichert P, 2002 Practical identifiability of asm2d parameters–systematic selection and tuning of parameter subsets. Water Res 36, 4113–4127. [PubMed: 12405420]

Burattini MN, Chen M, Chow A, Coutinho FAB, Goh KT, Lopez LF, M.A S, Massad E, 2008 Modelling the control strategies against dengue in Singapore. Epidemiol. Infect 136.

Centers for Disease Control Taiwan, Taiwan National Infectious Disease Statistics System, https://nidss.cdc.gov.tw/en/.

Chan M, Johansson MA, 2012 The incubation periods of dengue viruses. PLoS ONE 7, 1–7.

Chang MS, Christophel EM, Gopinath D, Abdur RM, 2011 Challenges and future perspective for dengue vector control in the Western Pacific Region. West. Pac. Surveill. Response 2, e1.

Chang SF, Huang JH, Shu PY, 2012 Characteristics of dengue epidemics in Taiwan. J. Formos. Med. Assoc 111, 297–299. [PubMed: 22748618]

Chao DL, Halstead SB, Halloran ME, Longini IM, 2012 Controlling dengue with vaccines in Thailand. PLoS Negl. Trop. Dis 6.

Chen S-C, Hsieh M-H, 2012 Modeling the transmission dynamics of dengue fever: implications of temperature effects. Sci. Tot. Environ 431, 385–391.

Chiroleu F, Dumont Y, 2010 Vector control for the Chikungunya disease. Math. Biosci. Eng 7, 313–345. [PubMed: 20462292]

Chis O-T, Banga JR, Balsa-Canto E, 2011 Structural identifiability of systems biology models: a critical comparison of methods. PLoS ONE 6, e27755. [PubMed: 22132135]

Chowell G, Diaz-Dueñas P, Miller JC, Alcazar-Velazco A, Hyman JM, Fenimore PW, Castillo-Chavez C, 2007 Estimation of the reproduction number of dengue fever from spatial epidemic data. Math. Biosci 208, 571–589. [PubMed: 17303188]

Christofferson RC, Mores CN, Wearing HJ, 2016 Bridging the gap between ex-perimental data and model parameterization for chikungunya virus transmission predictions. J. Infect. Dis 214, S466–S470. [PubMed: 27920175]

Cintrón-Arias A, Banks HT, Capaldi A, Lloyd AL, 2009 A sensitivity matrix based methodology for inverse problem formulation. J. Inverse Ill-posed Probl 17, 545–564.

Cobelli C, DiStefano JJ, 1980 Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. Am. J. Physiol 239, R7–R24. [PubMed: 7396041]

Coutinho FB, Burattini M, Lopez L, Massad E, 2005 An approximate threshold condition for non-autonomous system: an application to a vector-borne infection. Math. Comput. Simul 70, 149–158.

Coutinhoa FAB, Burattinia MN, Lopeza LF, Massada E, 2006 Threshold conditions for a non-autonomous epidemic system describing the population dynamics of dengue. Bull. Math. Biol 68, 2263–2282. [PubMed: 16952019]

Dommar CJ, Lowe R, Robinson M, Rodó X, 2014 An agent-based model driven by tropical rainfall to understand the spatio-temporal heterogeneity of a chikungunya outbreak. Acta Trop 129, 61–73. [PubMed: 23958228]

Dumont Y, Chiroleu F, Domerg C, 2008 On a temporal model for the chikungunya disease: modeling, theory and numerics. Math. Biosci 213, 80–91. [PubMed: 18394655]

Eisenberg MC, Hayashi MAL, 2014 Determining identifiable parameter combinations using subset profiling. Math. Biosci 256, 116–126. [PubMed: 25173434]

Eisenberg M, Robertson S, Tien J, 2013 Identifiability and estimation of multiple transmission pathways in waterborne disease. J. Theor. Biol 84–102.

Enduri MK, Jolad S, 2014 Dynamics of Dengue with Human and Vector Mobility

Erickson RA, Presley SM, Allen LJ, Long KR, Cox SB, 2010 A dengue model with a dynamic *Aedes albopictus* vector population. Ecol. Modell 221, 2899–2908.

Evans ND, White LJ, Chapman MJ, Godfrey KR, Chappell MJ, 2005 The structural identifiability of the susceptible infected recovered model with seasonal forcing. Math. Biosci 194, 175–197. [PubMed: 15854675]

Ferguson NM, Cucunuba ZM, Dorigatti I, Nedjati-Gilani GL, Donnelly CA, Basanez M-G, Nouvellet P, Lessler J, 2016a Countering the Zika epidemic in Latin America. Science (80-) 353, 353–354.

Ferguson NM, Rodríguez-Barraquer I, Dorigatti I, Mier-y Teran-Romero L, Laydon DJ, Cummings DAT, 2016b Benefits and risks of the Sanofi-Pasteur dengue vaccine: modeling optimal deployment. Science (80-) 353, 1033–1036.

Focks DA, Barrera R, 2006 Dengue transmission dynamics: assessment and implications for control. Rep. Sci. Work. Gr. Meet. Dengue 92–108.

Garba S, Gumel A, Abu Bakar M, 2008 Backward bifurcations in dengue transmission dynamics. Math. Biosci 215, 11–25. [PubMed: 18573507]

Gubler DJ, 2011 Dengue, urbanization and globalization: the unholy trinity of the 21st century. Trop. Med. Health 39, S3–S11.

Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, Hunsperger E, Kroeger A, Margolis HS, Martínez E, Nathan MB, Pelegrino JL, Simmons C, Yoksan S, Peeling RW, 2010 Dengue: a continuing global threat. Nat. Rev. Microbiol 8, S7–S16. [PubMed: 21079655]

Hales S, de Wet N, Maindonald J, Woodward A, 2002 Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. Lancet 360, 830–834. [PubMed: 12243917]

Hales S, Edwards SJ, Kovats RS, 2003 Impacts on health of climate extremes. Clim. Chang. Hum. Heal. Risks Responses 79–102.

Heffernan JM, Smith RJ, Wahl LM, 2005 Perspectives on the basic reproductive ratio. J. R. Soc. Interface 2, 281–293. [PubMed: 16849186]

Isidoro C, Fachada N, Barata F, Rosa A, 2011 Agent-based model of dengue disease transmission by Aedes aegypti populations. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 5777 LNAI pp. 345–352.

Jacquez JA, Greif P, 1985 Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design. Math. Biosci 77, 201–227.

Johnson LR, Ben-Horin T, Lafferty KD, McNally A, Mordecai E, Paaijmans KP, Pawar S, Ryan SJ, 2015 Understanding uncertainty in temperature effects on vector-borne disease: a bayesian approach. Ecology 96, 203–213. [PubMed: 26236905]

Kearney M, Porter WP, Williams C, Ritchie S, Hoffmann AA, 2009 Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito Aedes aegypti in Australia. Funct. Ecol 23, 528–538.

Kermack WO, McKendrick a.G., 1927 Contributions to the mathematical theory of epidemics. Proc. R. Soc. Lond 115, 700–721.

Khan A, Hassan M, Imran M, 2014 Estimating the basic reproduction number for single-strain dengue fever epidemics. Infect. Dis. Poverty 3, 12. [PubMed: 24708869]

Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Elyazar IR, Teng H-J, Brady OJ, Messina JP, Pigott DM, Scott TW, Smith DL, Wint GW, Golding N, Hay SI, 2015 The global distribution of the arbovirus vectors *Aedes aegypti* and Ae. albopictus. Elife 4.

Kucharski AJ, Funk S, Eggo RM, Mallet H-P, Edmunds WJ, Nilles EJ, 2016 Transmission dynamics of zika virus in island populations: a modelling analysis of the 2013–14 French polynesia outbreak. PLoS Negl. Trop. Dis 10, e0004726. [PubMed: 27186984]

Laneri K, Bhadra A, Ionides EL, Bouma M, Dhiman RC, Yadav RS, Pascual M, 2010 Forcing versus feedback: epidemic malaria and monsoon rains in Northwest India. PLoS Comput. Biol 6.

Li J, Zou X, 2009 Modeling spatial spread of infectious diseases with a fixed latent period in a spatially continuous domain. Bull. Math. Biol 71, 2048–2079. [PubMed: 19787405]

Manore CA, Hickmann KS, Xu S, Wearing HJ, Hyman JM, 2014 Comparing dengue and chikungunya emergence and endemic transmission in A. aegypti and A. albopictus. J. Theor. Biol 356, 174–191. [PubMed: 24801860]

Manore CA, Hickmann KS, Hyman JM, Foppa IM, Davis JK, Wesson DM, Mores CN, 2015 A network-patch methodology for adapting agent-based models for directly transmitted disease to mosquito-borne disease. J. Biol. Dyn 9, 52–72. [PubMed: 25648061]

McLennan-Smith TA, Mercer GN, 2014 Complex behaviour in a dengue model with a seasonally varying vector population. Math. Biosci 248, 22–30. [PubMed: 24291301]

Mendes Luz P, Torres Codeço C, Massad E, Struchiner CJ, 2003 Uncertainties regarding dengue modeling in Rio de Janeiro, Brazil. Mem. Inst. Oswaldo Cruz 98, 871–878. [PubMed: 14765541]

Meshkat N, Anderson C, DiStefano III JJ, 2012 Alternative to Ritt's pseudodivision for finding the input–output equations of multi-output models. Math. Biosci 239, 117–123. [PubMed: 22626896]

Meshkat N, Kuo CE-Z, DiStefano J, 2014 On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: a novel web implementation. PLOS ONE 9, e110261. [PubMed: 25350289]

Miao H, Xia X, Perelson AS, Wu H, 2011 On identifiability of nonlinear ODE models and applications in viral dynamics. SIAM Rev 53, 3–39.

Mordecai EA, Paaijmans KP, Johnson LR, Balzer C, Ben-Horin T, de Moor E, McNally A, Pawar S, Ryan SJ, Smith TC, Lafferty KD, 2013 Optimal temperature for malaria transmission is dramatically lower than previously predicted. Ecol. Lett 16, 22–30. [PubMed: 23050931]

Mordecai EA, Cohen JM, Evans MV, Gudapati P, Johnson LR, Lippi CA, Miazgowicz K, Murdock CC, Rohr JR, Ryan SJ, Savage V, Shocket MS, Stewart Ibarra A, Thomas MB, Weikel DP, 2017 Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. PLoS Negl. Trop. Dis 11, e0005568. [PubMed: 28448507]

Morrison AC, Zielinski-Gutierrez E, Scott TW, Rosenberg R, 2008 Defining challenges and proposing solutions for control of the virus vector *Aedes aegypti*. PLoS Med 5, e68. [PubMed: 18351798]

Morrison AC, Ellis AM, Garcia AJ, Scott TW, Focks DA, 2011 Parameterization and sensitivity analysis of a complex simulation model for mosquito population dynamics, dengue transmission, and their control. Am. J. Trop. Med. Hyg 85, 257–264. [PubMed: 21813844]

Moulay D, Aziz-Alaoui M, Cadivel M, 2011 The chikungunya disease: modeling, vector and transmission global dynamics. Math. Biosci 229, 50–63. [PubMed: 21070789]

Moulay D, Aziz-Alaoui MA, Kwon H-D, 2012a Optimal control of chikungunya disease: Larvae reduction, treatment and prevention. Math. Biosci. Eng 9, 369–392. [PubMed: 22901069]

Moulay D, Verdière N, Denis-Vidal L, 2012b Identifiablility of Parameters in an Epidemiologic Model Modeling the Transmission of the Chikungunya

Musso D, Cao-Lormeau VM, Gubler DJ, 2015 Zika virus: following the path of dengue and chikungunya? Lancet 386, 243–244. [PubMed: 26194519]

Newton EAC, Reiter P, 1992 A model of the transmission of dengue fever with an evaluation of the impact of ultra-low volume (ULV) insecticide applications on dengue epidemics. Am. J. Trop. Med. Hyg 47, 709–720. [PubMed: 1361721]

Oki M, Sunahara T, Hashizume M, Yamamoto T, 2011 Optimal timing of insecticide fogging to minimize dengue cases: modeling dengue transmission among various seasonalities and transmission intensities. PLoS Negl. Trop. Dis 5, e1367. [PubMed: 22039560]

Ollivier F, 1990 Le probleme de l'identifiabilite structurelle globale: approche theorique, methodes effectives et bornes de complexite. École Polytechnique Ph.D. Thesis

Pandey A, Mubayi A, Medlock J, 2013 Comparing vector-host and SIR models for dengue transmission. Math. Biosci 246, 252–259. [PubMed: 24427785]

Patz JA, Martens WJ, Focks DA, Jetten TH, 1998 Dengue fever epidemic potential as projected by general circulation models of global climate change. Environ. Health Perspect 106, 147–153. [PubMed: 9452414]

Perkins A, Siraj A, Ruktanonchai WC, Kraemer M, Tatem A, 2016 Model-based projections of Zika virus infections in childbearing women in the Americas. bioRxiv 1, 039610.

Pia Saccomani M, Audoly S, Bellu G, D'Angio L, 2001 A new differential algebra algorithm to test identifiability of nonlinear systems with given initial conditions. In: Proc. 40th IEEE Conf. Decis. Control (Cat. No.01CH37228). IEEE. pp. 3108–3113.

Pinho STR, Ferreira CP, Esteva L, Barreto FR, Morato e Silva VC, Teixeira MGL, 2010 Modelling the dynamics of dengue real epidemics. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci 368, 5679–5693.

Poletti P, Messeri G, Ajelli M, Vallorani R, Rizzo C, Merler S, 2011 Transmission potential of chikungunya virus and control measures: the case of Italy. PLoS ONE 6.

Powell JR, Tabachnick WJ, 2013 History of domestication and spread of *Aedes aegypti* – a review. Mem. Inst. Oswaldo Cruz 108, 11–17. [PubMed: 24473798]

Prosper O, Ruktanonchai N, Martcheva M, 2012 Assessing the role of spatial heterogeneity and human movement in malaria dynamics and control. J. Theor. Biol 303, 1–14. [PubMed: 22525434]

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J, 2009 Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923–1929. [PubMed: 19505944]

Reich NG, Shrestha S, King AA, Rohani P, Lessler J, Kalayanarooj S, Yoon I-K, Gibbons RV, Burke DS, Cummings DA, 2013 Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. J. R. Soc. Interface 10, 20130414. [PubMed: 23825116]

Reiner RC, Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM, George DB, Le Menach A, Pulliam JRC, Bisanzio D, Buckee C, Chiyaka C, Cummings DAT, Garcia AJ, Gatton ML, Gething PW, Hartley DM, Johnston G, Klein EY, Michael E, Lindsay SW, Lloyd AL, Pigott DM, Reisen WK, Ruktanonchai N, Singh BK, Tatem AJ, Kitron U, Hay SI, Scott TW, Smith DL, 2013 A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. J. R. Soc. Interface 10, 20120921. [PubMed: 23407571]

Reiner RC, Stoddard ST, Forshey BM, King AA, Ellis AM, Lloyd AL, Long KC, Rocha C, Vilcarromero S, Astete H, Bazan I, Lenhart A, Vazquez-Prokopec GM, Paz-Soldan VA, McCall PJ, Kitron U, Elder JP, Halsey ES, Morrison AC, Kochel TJ, Scott TW, 2014 Time-varying, serotype-specific force of infection of dengue virus. Proc. Natl. Acad. Sci. U. S. A 111, E2694–E2702. [PubMed: 24847073]

Rothenberg TJ, 1971 Identification in parametric models. Econom. J. Econom. Soc 39, 577–591.

Rudolph KE, Lessler J, Moloney RM, Kmush B, Cummings DAT, 2014 Review Article: Incubation Periods of Mosquito-Borne Viral Infections: A Systematic Review

Sardar T, Sasmal SK, Chattopadhyay J, 2016 Estimating dengue type reproduction numbers for two provinces of Sri Lanka during the period 2013–14. Virulence 7, 187–200. [PubMed: 26646355]

Scott TW, Morrison AC, 2003 Aedes aegypti density and the risk of dengue virus transmission. Ecol. Asp. Appl. Genet. Modif. Mosq 187–206.

Shutt DP, Manore CA, Pankavich S, Porter AT, Del Valle SY, 2017 Estimating the reproductive number, total outbreak size, and reporting rates for Zika epidemics in South and Central America. Epidemics

Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE, 2012 Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens

Tuncer N, Le TT, 2018 Structural and practical identifiability analysis of outbreak models. Math. Biosci 299, 1–18. [PubMed: 29477671]

Tuncer N, Gulbudak H, Cannataro VL, Martcheva M, 2016 Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever. Bull. Math. Biol 78, 1796–1827. [PubMed: 27651156]

van den Driessche P, Watmough J, 2002 Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci 180, 29–48. [PubMed: 12387915]

Weaver SC, 2013 Urbanization and geographic expansion of zoonotic arboviral diseases: mechanisms and potential strategies for prevention. Trends Microbiol 21, 360–363. [PubMed: 23910545]

WHO-VMI Dengue Vaccine Modeling Group, 2012 Assessing the potential of a candidate dengue vaccine with mathematical modeling. PLoS Negl. Trop. Dis 6, e1450. [PubMed: 22479655]

Wilder-Smith A, Gubler DJ, 2008 Geographic expansion of dengue: the impact of international travel. Med. Clin. N. Am 92, 1377–1390. [PubMed: 19061757]

Wilder-Smith A, 2012 Dengue infections in travellers. Paediatr. Int. Child Health 32, 28–32. [PubMed: 22668447]

World Health Organization, 2009 Research Special Programme for Diseases and Training in Tropical, Dengue: guidelines for diagnosis, treatment, prevention, and control. Dengue Guidel. Diagn. Treat. Prev. Control 160.

World Health Organization, 2011 Report of the Meeting of the WHO/VMI Workshop on Dengue Modeling, Technical Report World Health Organization, Geneva, Switzerland.

World Health Organization, 2012 Global Strategy for Dengue Prevention and Control 2012–2020, Technical Report

World Health Organization, 2017 Dengue and Severe Dengue http://www.who.int/mediacentre/factsheets/fs117/en/.

Wu H, Wang C, Teng H, Lin C, Lu L, Jian S, Chang N, Wen T, Wu J, Liu D, Lin L, Norris DE, Wu H, 2013 A dengue vector surveillance by human population-stratified ovitrap survey for Aedes (Diptera: Culicidae) adult and egg collections in high dengue-risk areas of Taiwan. J. Med. Entomol 50, 261–269. [PubMed: 23540112]

Yakob L, Clements ACA, 2013 A mathematical model of chikungunya dynamics and control: the major epidemic on Réunion Island. PLOS ONE 8.

Yang HM, Ferreira CP, 2008 Assessing the effects of vector control on dengue transmission. Appl. Math. Comput 198, 401–413.

Yang HM, Macoris MLG, Galvani KC, Andrighetti MTM, Wanderley DMV, 2009a Assessing the effects of temperature on dengue transmission. Epidemiol. Infect 137, 1179. [PubMed: 19192323]

Yang HM, Macoris MLG, Galvani KC, Andrighetti MTM, Wanderley DMV, 2009b Assessing the effects of temperature on the population of *Aedes aegypti*, the vector of dengue. Epidemiol. Infect 137, 1188–1202. [PubMed: 19192322]

Zhu S, Lilianne D-V, Verdière N, 2015 Identifiability Study in a Model Describing the Propagation of the Chikungunya to the Human Population

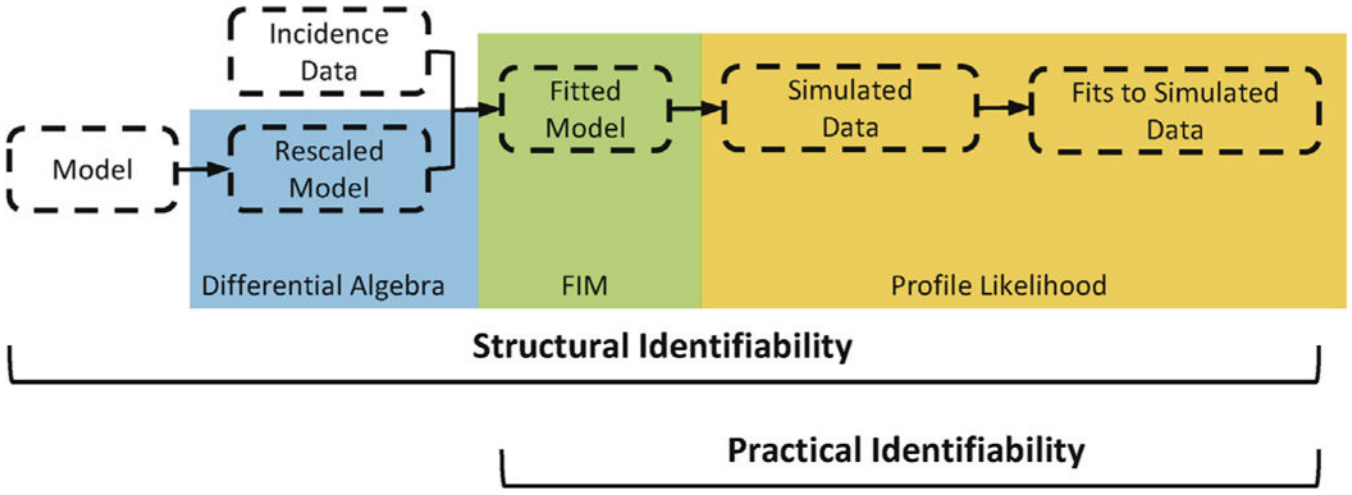

Fig. 1.
Summary of the parameter estimation and identifiability analysis process.

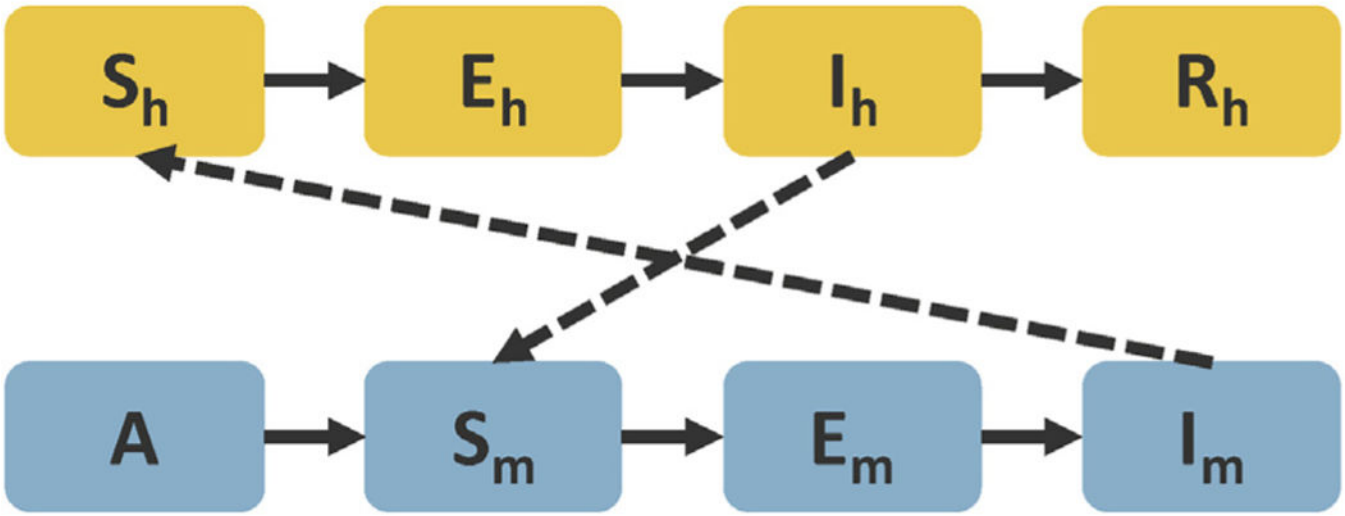

Fig. 2.

Diagram of the SEIR-based model. Subscript *h* indicates human, *m* indicates mosquitoes, and *S, E, I, R* represent susceptible, latent (exposed), infectious, and recovered humans or adult mosquitoes. A represents immature mosquitoes (larvae and pupae).

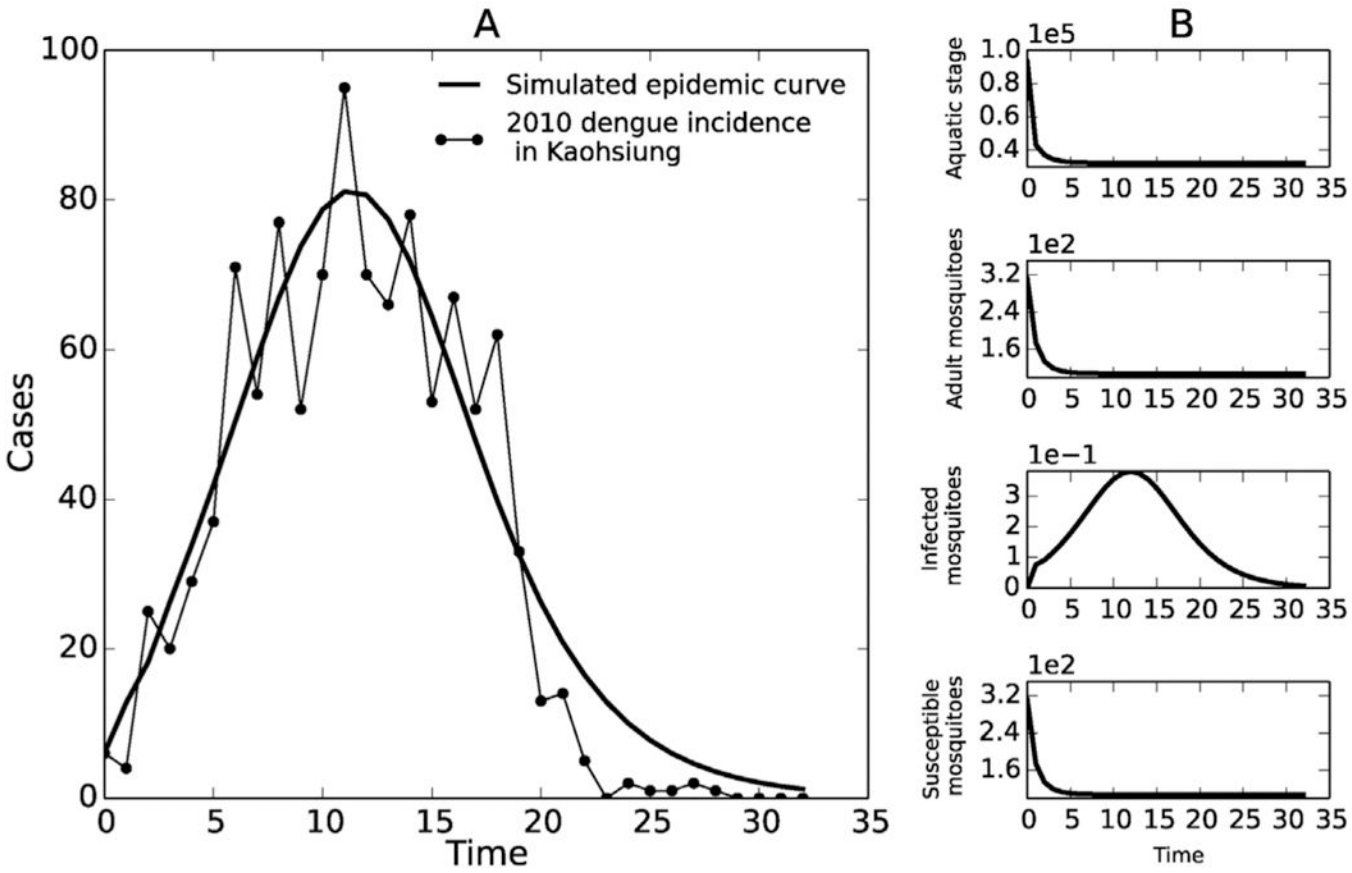

Fig. 3.

(A) model (dotted line) fitted to weekly incidence data (black circles) in Kaohsiung, Taiwan (2010); (B) simulated mosquito population data corresponding to Scenarios 2–4.

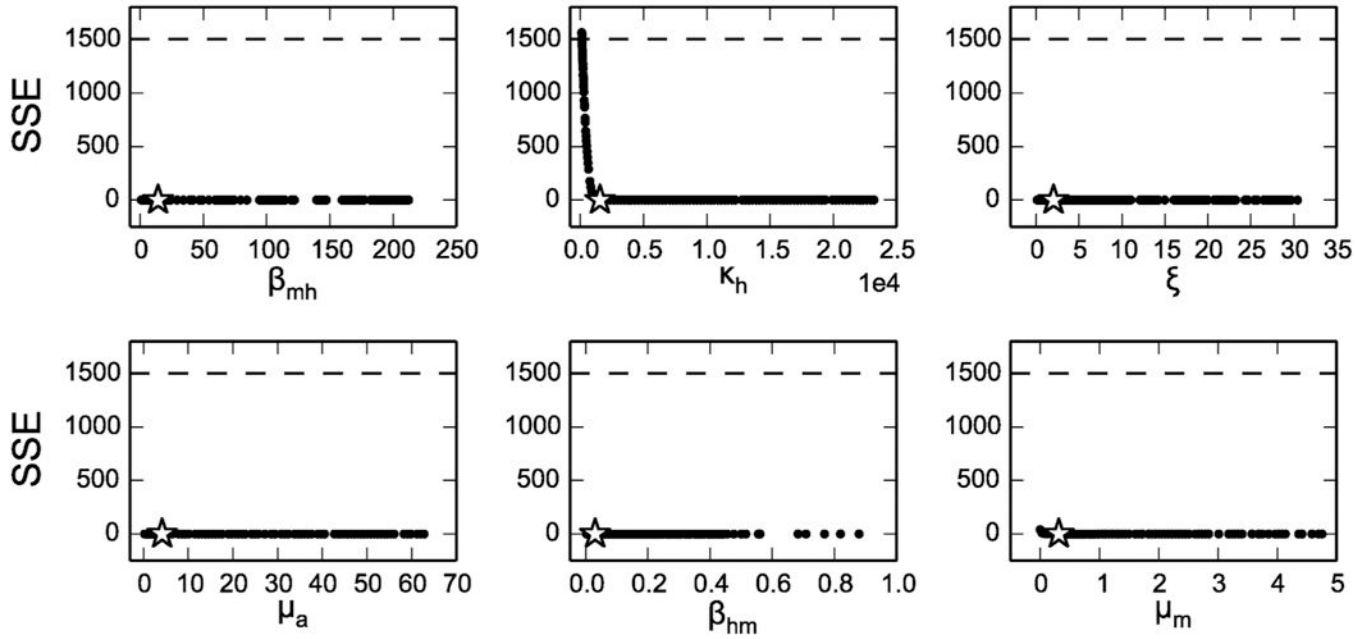

Fig. 4.

Profile likelihoods (black circles) assuming simulated, noise-free human incidence data (Scenario 1). Stars indicate the minimum sum of squared error (SSE) and dashed lines indicate the threshold for 95% confidence bounds. All six fitted model parameters are practically unidentifiable, with shallow minima which do not cross the confidence interval threshold within realistic biological ranges (zoomed in versions of the profiles showing the minima are given in the Supplementary Information).

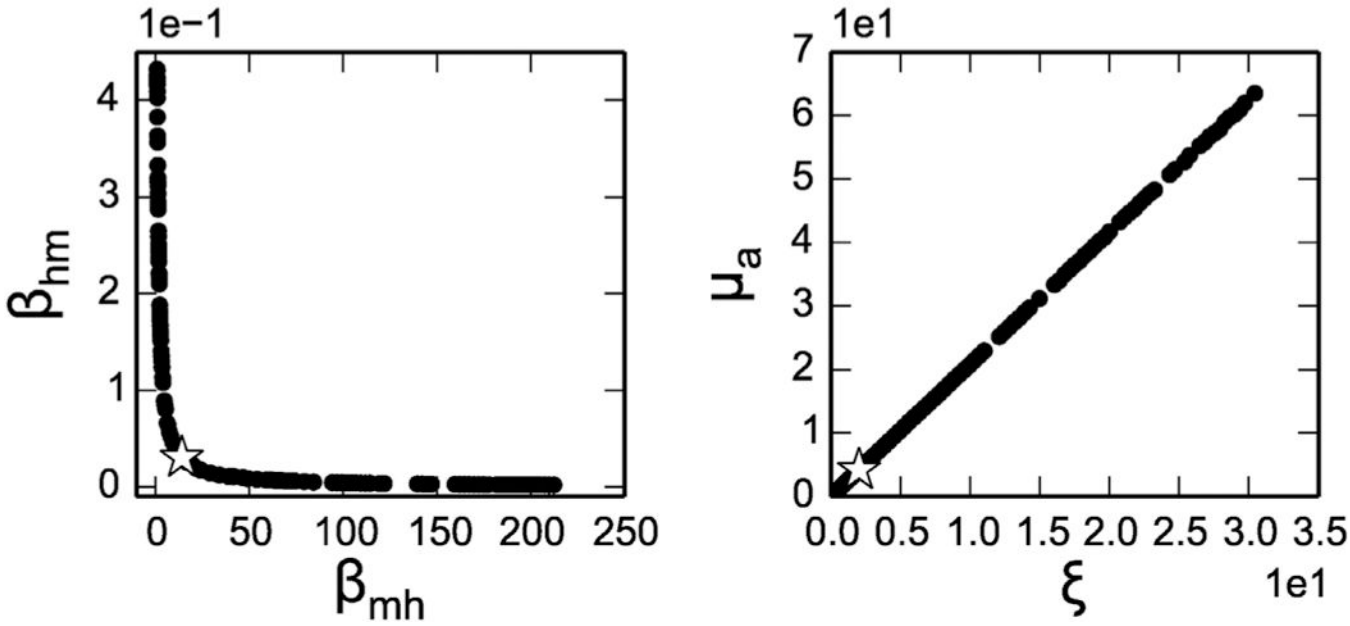

Fig. 5.

Parameter relationship scatter plots derived from the Scenario 1 profiles shown in Fig. 4, showing the relationships between β_{hm} and β_{mh} as β_{mh} is profiled and between ξ and μ_a as ξ is profiled. The two parameters in each pair compensate for one another, leading to the flat profile observed in Fig. 4.

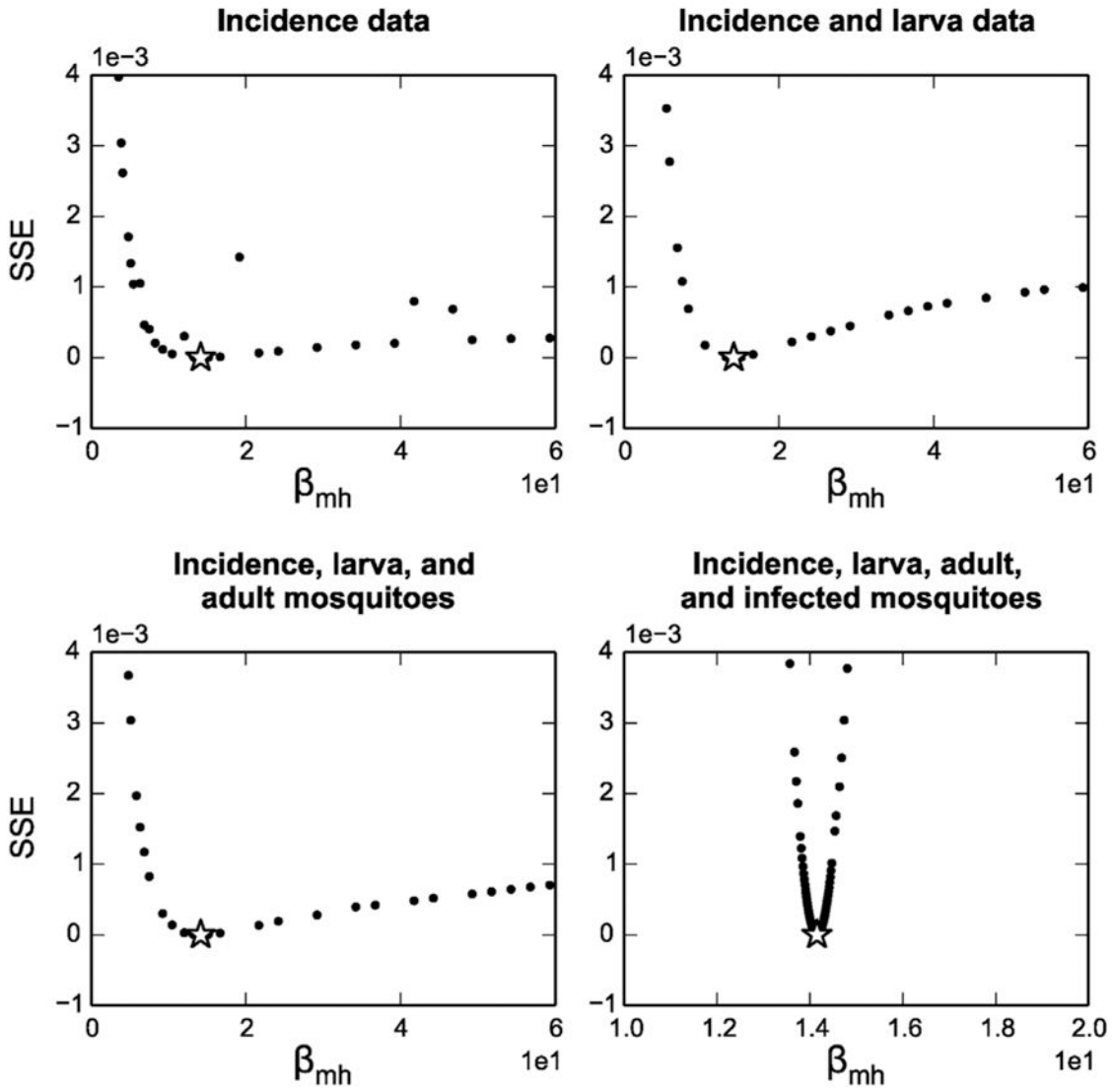

Fig. 6.

Profile likelihoods for β_{mh} with human incidence data only (Scenario 1), human incidence and larva count data (Scenario 2), human incidence, larva counts, and adult mosquito counts (Scenario 3), and data for human incidence, larva counts, adult mosquito counts, and infected adult mosquito counts (Scenario 4).

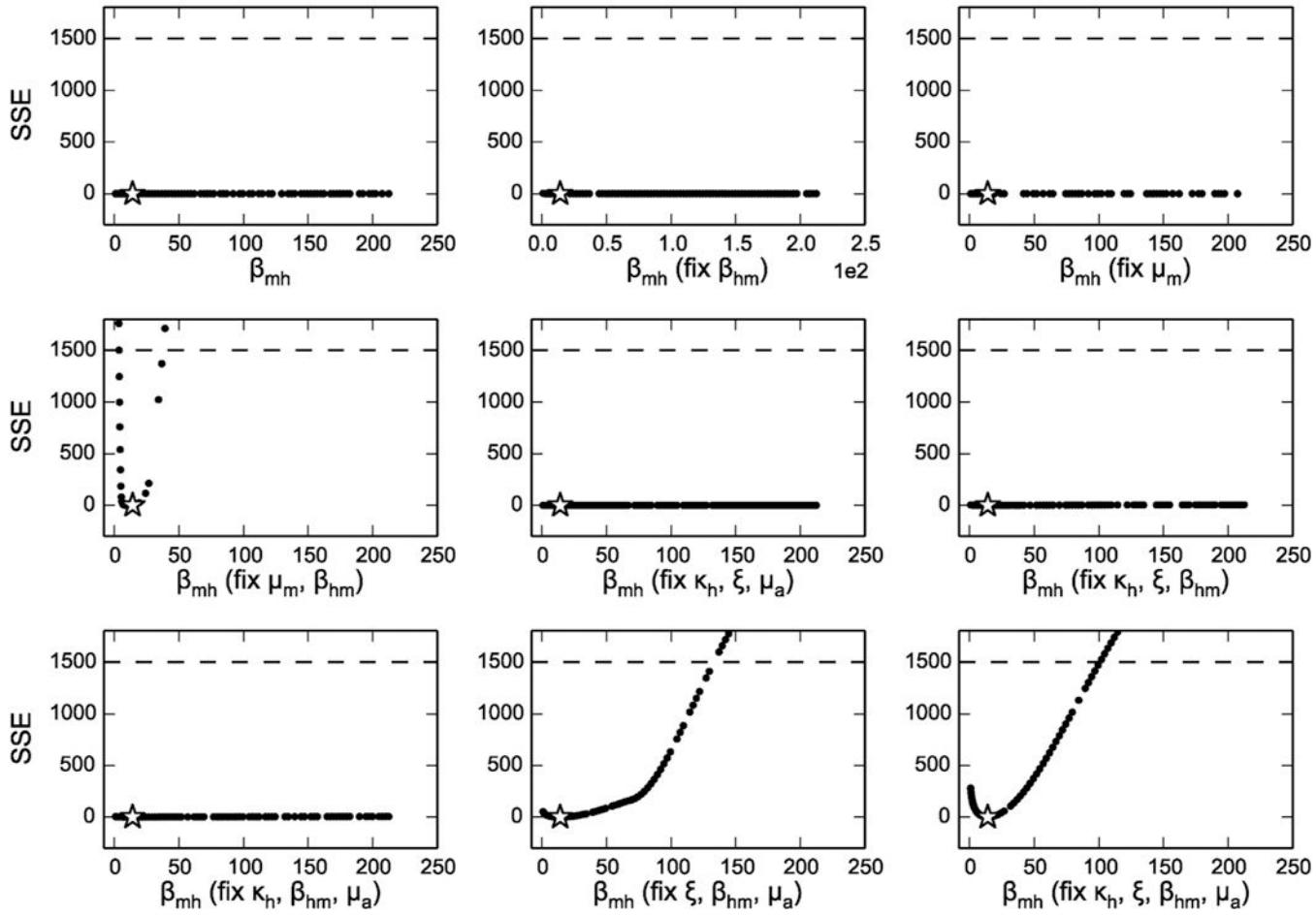

Fig. 7.

Profile likelihoods of β_{mh} when only subsets of μ_a, ξ, κ, μ_m and β_{mh} are fitted. The fixed subset (in addition to β_{mh}) is shown in parentheses on the x-axis.

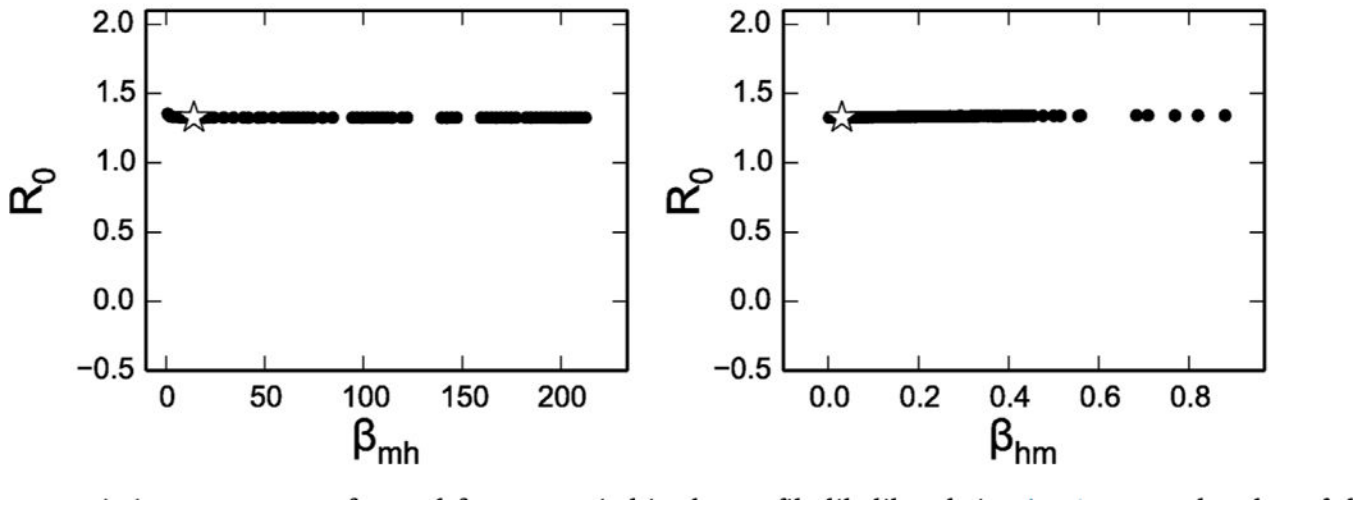

Fig. 8.

Values for $\mathcal{R}_0$ as the two transmission parameters, β_{mh} and β_{hm} are varied in the profile likelihoods in Fig. 4. For each value of the profiled parameter, the plotted $\mathcal{R}_0$ value uses to the best-fit values of the remaining parameters. $\mathcal{R}_0$ remains relatively constant over the profiled parameter range, in spite of large changes in the parameter values.

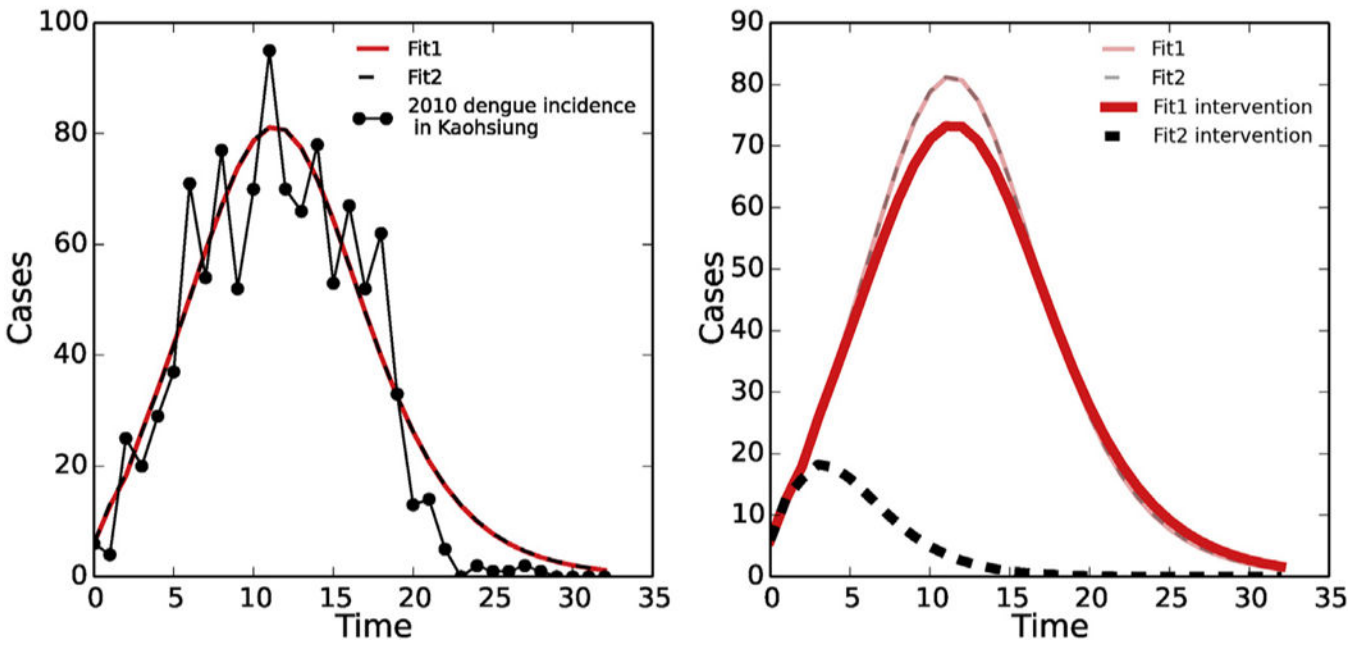

Fig. 9.

Illustration of the implications of model unidentifiability on intervention prediction. Left: two model simulations using different parameter values that give the same fit to data, based on the profiles in Fig. 4 (red solid line – original fitted parameter values from Table 1; black dashed line – [$\beta_{mh} = 38.10$, $\kappa_h = 1625.42$, $\xi = 0.13$, $\mu_a = 0.15$, $\beta_{hm} = 0.02$, $\mu_m = 0.42$]). Right: Simulated intervention results for both parameter sets, supposing that 10% of the aquatic (immature) mosquito population is removed at each time step.

Table 1

Parameter estimates and values. Estimated parameters are marked in bold; confidence bounds and uncertainty for the fitted parameters are examined further below.

Parameter	Description (units)	Value	Source
μ	Human birth and death rate (day^{-1})	0	Newton and Reiter (1992)
$\boldsymbol{\beta_{mh}}$	Rescaled mosquito-to-human infection rate (day^{-2})	14.15	Fitted
α	Intrinsic incubation rate (day^{-1})	0.14	World Health Organization (2009), Chan and Johansson (2012), Rudolph et al. (2014)
η	Recovery rate (day^{-1})	0.2	World Health Organization (2009)
$\boldsymbol{\xi}$	Rescaled oviposition-fertilization rate of larvae (day^{-2})	2.03	Fitted
$\boldsymbol{\mu_a}$	Additionaldeathrateforaquatic(immature)mosquitoesduringtheepidemic,due to interventions and environmental changes (day^{-1})	4.18	Fitted
$\boldsymbol{\beta_{hm}}$	Human-to-mosquito infection rate (day^{-1})	0.03	Fitted
$\boldsymbol{\mu_m}$	Mosquito death rate (day^{-1})	0.32	Fitted
γ	Extrinsic incubation rate (day^{-1})	0.1	World Health Organization (2009), Chan and Johansson (2012), Rudolph et al. (2014)
$\boldsymbol{\kappa_h}$	Fraction of cases reported multiplied by total human population at risk (number of individuals)	1546.74	Fitted
κ_a	Maximum possible immature mosquito counts observed in traps: fraction of aquatic mosquitoes observed times total maximal carrying capacity of aquatic mosquitoes (used for simulated data only) (number of mosquitoes)	93,420	Wu et al. (2013)
κ_m	Maximum possible observed growth rate of new adult mosquitoes: fraction of adults mosquitoes observed times the maximum maturation rate of mosquitoes (used for simulated data only) (number of mosquitoes/day)	98.71	Wu et al. (2013)