Epidemics 26 (2019) 43–57

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

# Dengue modeling in rural Cambodia: Statistical performance versus epidemiological relevance

Clara Champagne[a,b,*], Richard Paul[c,d], Sowath Ly[e], Veasna Duong[f], Rithea Leang[g], Bernard Cazelles[a,d,h,*]

[a] Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Research University, 75005 Paris, France
[b] CREST, ENSAE, Université Paris Saclay, 5 avenue Henry Le Chatelier, 91120 Palaiseau, France
[c] Institut Pasteur, Unité de Génétique Fonctionnelle des Maladies Infectieuses, Department of Genomes and Genetics, F-75724 Paris cedex 15, France
[d] Centre National de la Recherche Scientifique (CNRS), Génomique évolutive, modélisation et santé, UMR 2000, 75724 Paris cedex 15, France
[e] Institut Pasteur in Cambodia, Epidemiology and Public Health Unit, Phnom Penh, Cambodia
[f] Institut Pasteur in Cambodia, Virology Unit, Phnom Penh, Cambodia
[g] National Dengue Control Program, Cambodia
[h] International Center for Mathematical and Computational Modeling of Complex Systems (UMMISCO), UMI 209 UPMC/IRD, Bondy cedex, France

ARTICLE INFO

*Keywords:*
Model selection
Dengue
Vector borne disease
Mathematical model
Cambodia

ABSTRACT

Dengue dynamics are shaped by the complex interplay between several factors, including vector seasonality, interaction between four virus serotypes, and inapparent infections. However, paucity or quality of data do not allow for all of these to be taken into account in mathematical models. In order to explore separately the importance of these factors in models, we combined surveillance data with a local-scale cluster study in the rural province of Kampong Cham (Cambodia), in which serotypes and asymptomatic infections were documented. We formulate several mechanistic models, each one relying on a different set of hypotheses, such as explicit vector dynamics, transmission via asymptomatic infections and coexistence of several virus serotypes. Models are confronted with the observed time series using Bayesian inference, through Markov chain Monte Carlo. Model selection is then performed using statistical information criteria, and the coherence of epidemiological characteristics (reproduction numbers, incidence proportion, dynamics of the susceptible classes) is assessed in each model. Our analyses on transmission dynamics in a rural endemic setting highlight that two-strain models with interacting effects better reproduce the long term data, but they are difficult to parameterize when relying on incidence cases only. On the other hand, considering the available data, incorporating vector and asymptomatic components seems of limited added-value when seasonality and underreporting are already accounted for.

## 1. Introduction

Dengue is a vector-borne viral disease transmitted by *Aedes* spp. caused by any of four dengue virus (DENV) serotypes. Infection can result in a flu-like illness, and sometimes potentially lethal complications called Dengue Hemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS), although a significant proportion are subclinical or asymptomatic, causing insufficient discomfort for clinical presentation (Grange et al., 2014). Dengue is ubiquitous in the tropics and the subtropics, particularly in Southeast Asia, the Pacific and the Americas (Guzman et al., 2010). The World Health Organization (WHO) considers that dengue is a major public health issue worldwide, with four billion people in 128 countries exposed to the dengue virus (Messina et al., 2013; WHO, 2017), an estimated 390 million infections every year and about 50–100 million symptomatic cases worldwide and a high disease burden (Bhatt et al., 2013; Shepard et al., 2016). Nowadays, there are more cases of dengue worldwide than any other arboviral disease (Halstead, 2007; Messina et al., 2015; Sharp et al., 2017).

The value of mathematical models and associated statistical tools for investigating public health policy questions has long been recognized and has provided insights into their transmission and control for more than one hundred years (Heesterbeek et al., 2015; Reiner et al., 2013). It is important, however, to adapt them as much as possible to a specific setting, in order to derive appropriate public health recommendations and accurately generate the key parameters using estimation tools, so that they can produce realistic conclusions, in

* Corresponding authors at: Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Research University, 75005 Paris, France.
*E-mail addresses:* champagn@biologie.ens.fr (C. Champagne), cazelles@biologie.ens.fr (B. Cazelles).

https://doi.org/10.1016/j.epidem.2018.08.004
Received 6 November 2017; Received in revised form 19 July 2018; Accepted 27 August 2018
Available online 29 August 2018

**Table 1**
**Number of cases under 15 years old in the DENFREE study (index cases and community cases) and in the surveillance system NDSS, and associated theoretical quantities in models.** One case in 2013 was coinfected with DENV-1 and DENV-2 and was not included in models with two strains (serotypes).

| Children under 15 years old | Total number of observed cases | | Theoretical quantity | Observation rate | Model |
|---|---|---|---|---|---|
| DENFREE data | year 2012 | year 2013 | | | |
| Confirmed symptomatic cases | 236 | 574 | $C_I$ | $r_D$ | All |
| Denv-1 | 226 | 451 | $C_{I1}$ | $r_D$ | SEIR2, SEIR2psi |
| Denv-2, Denv-3, Denv-4 | 5 | 122 | $C_{I2}$ | $r_D$ | SEIR2, SEIR2psi |
| Unknown serotype | 5 | 0 | | | |
| Confirmed asymptomatic cases | 5 | 28 | $C_A$ | $r_A$ | SEIAR |
| Denv-1 | 5 | 21 | | | |
| Denv-2, Denv-3, Denv-4 | 0 | 7 | | | |
| NDSS data | 2002–2015 | | | | |
| Surveillance cases | 10,780 (10,096 in 2002–2013) | | $C_I$ | $r_N$ | All except SEIAR |
| | | | $C_H$ | $r_H$ | SEIAR |

accordance with the observed data.

Dengue dynamics are shaped by the complex interplay between many factors associated with the mosquito vector and human hosts and their interactions with the virus. Hitherto, the exploration of dengue dynamics has focused on the urban setting, where the incidence of dengue is highest (Clapham et al., 2015; Nisalak et al., 2016; Reich et al., 2013; Salje et al., 2012). Few studies have been carried out in rural settings (Aldstadt et al., 2012; Mammen et al., 2008; Strickman et al., 2000), despite growing evidence that rural dengue is an increasing problem. Guha-Sapir and Schimmer (2005) observed shifts in modal age, rural spread, and social determinants of dengue susceptibility, with major implications for health services. Muhammad Azami et al. (2011) observed similar dengue seroprevalence rates between urban and rural samples, showing that dengue is not confined to urban areas in Malaysia. Chareonsook et al. (1999) showed that DHF in Thailand, which was originally thought to be an urban disease, has spread to most areas of Thailand, and is now more common in rural than urban areas and studies suggest that rural dengue incidence can surpass urban and semi-urban communities within the same region (Reller et al., 2012; Vong et al., 2010). In addition, several studies have stressed that rural settings play an important role in the timing of dengue epidemics in Southeast Asia, with the seasonal dengue waves typically arriving later in major urban centers (Cazelles and Cazelles, 2014; Cuong et al., 2013; Teurlai et al., 2012).

In this study, we combine two datasets from rural Cambodia that provide information on different key factors. We contrast and compare several mechanistic models, incorporating differing levels of complexity with respect to vector dynamics, coexistence of several virus strains, and transmission via asymptomatic infections. Models are adapted to the observed time series using Bayesian inference, through Markov Chain Monte Carlo (MCMC) and compared in light of the data, using statistical indicators to identify the best model (Camacho et al., 2011; King et al., 2008; Pandey et al., 2013; Reich et al., 2013). In addition, we also analyze the epidemiological coherence of the estimated models in simulations. Critically, we do not merely focus on the observed infected individuals but also on other compartments, such as the susceptible class of individuals. By comparing these models, we try to find a realistic but parsimonious way of modeling dengue epidemics in rural Cambodia. The best model may then be used in the study of intervention scenarios or in comparative analyses with other settings. For instance, it could be readily expanded to understand the potential impact of different vaccination strategies in rural settings.

## 2. Methods

### 2.1. Data

#### 2.1.1. Study area

Kampong Cham province is a densely populated rural province 120 km northeast from the capital Phnom Penh. Dengue is endemic and strongly seasonal, with outbreaks occurring every year from June to September, during the rainy season. The four virus serotypes co-circulate, even though one usually dominates the three others for about 3 to 5 years. We used two different datasets reporting dengue cases in the province: the results of a punctual study conducted in a 30 km radius around the city of Kampong Cham (DENFREE data), and the national surveillance data (NDSS data) in the four districts comprising the DENFREE study area (Kampong Cham, Kampong Siem, Prey Chhor and Tboung Khmum, with the administrative divisions of 2012–2013).

#### 2.1.2. DENFREE data

The DENFREE study took place in the Kampong Cham region during the 2012 and 2013 outbreaks. Patients with acute dengue-like illness were enrolled in three hospitals in the Kampong Cham province. Positive DENV cases were considered as index cases, and an outbreak investigation was initiated in their neighbourhood, in order to detect additional symptomatic cases but also asymptomatic or mildly symptomatic cases. For both index and outbreak investigation cases, DENV infection was confirmed by qRT-PCR. The study protocol is extensively detailed in Duong et al. (2015).

We used the series of the total number of cases per week (index cases and outbreak investigation cases) between 18th June–5th November 2012 and 3rd June–30th September 2013 (Table 1). We also restricted the study to children under 15 years old for two major reasons: most of the reported dengue cases were in this age class (89.6% and 90.8% in 2012 and 2013 respectively, cf. Appendix B), and it allowed a comparison with other dengue reporting systems in Cambodia, which are mainly done at paediatric hospitals. Information on the serotype responsible for infection, and symptomatic/asymptomatic status of the patients were available (Table 2).

#### 2.1.3. NDSS data

Because the DENFREE data covers only a relatively short period of time, surveillance data were added to improve the estimations. Surveillance of dengue is conducted at the national level in Cambodia, through the National Dengue Surveillance system (NDSS) (Huy et al., 2010; Teurlai et al., 2012), involving the paediatric departments of several hospitals throughout the country. Since surveillance is hospital based, mostly severe cases are observed. Diagnosis is done clinically and only a small fraction of the cases are confirmed serologically. Because of the co-circulation of other flaviviruses (Chikungunya, Japanese Encephalitis) and the relative non-specificity of symptoms, clinical misdiagnosis may be frequent. Since surveillance is carried out in paediatric departments, only cases among children under 16 years old are reported.

We selected all the cases under 15 years old in the four districts involved in the DENFREE study between January 2002 and December 2015 and aggregated them per week (cases under 15 years old

**Table 2**
Comparison of both datasets.

| | DENFREE | NDSS |
|---|---|---|
| Time window | Observations in 2012–2013 | Observations in 2002–2015 |
| | No observation during inter-epidemic period, dataset starts during the epidemic peak | Observations all year round |
| Area | 30 km radius around Kampong Cham city | 4 districts comprising the DENFREE study |
| | Clustered collecting process | Stable reporting process over time |
| | An observation rate can be calculated | Unknown observation rate |
| Diagnosis | Laboratory confirmation of all cases (qRT-PCR) | Clinical diagnosis |
| | Observation of symptomatic infections (index cases and outbreak investigation) and mildly symptomatic or asymptomatic infections (outbreak investigation) | Symptoms ranging from dengue fever to DHF and DSS (mostly hospitalized) |
| | Known serotype for both index and outbreak investigation cases | Unknown serotype |
| | Primary and secondary infections are not distinguished | Primary and secondary infections are not distinguished |

represent 96.2% of reported cases over the period, cf. Appendix B). In this area, on average, 770 cases under 15 years old are reported per year (maximum 1985 cases in 2007, minimum 209 cases in 2014). We used data from 2002 to 2013 for estimations, and data for 2014 and 2015 as the test set.

#### 2.1.4. Population

We take as the reference population ($N = 161, 391$) the number of children below 15 years old in four districts of the Kampong Cham province (Kampong Cham, Kampong Siem, Prey Chhor and Tboung Khmum, with the administrative divisions of 2012–2013) according to 2008 National Census (National Institute of Statistics, Ministry of Planning, 2009). Since the DENFREE study was conducted in a subpart of this area, we calculated the total population for the DENFREE study ($n = 65, 208$) as the sum of the population of children under 15 years old in all the villages investigated in either 2012 or 2013 (National Institute of Statistics, Ministry of Planning, 2009).

### 2.2. Models

All model parameters are defined in the figures captions and in Table 3.

#### 2.2.1. One-strain models

We take a Susceptible-Exposed-Infected-Recovered (SEIR) model as the simplest model (cf. Fig. 1). $C_I$ represents the count of new cases and it is aggregated weekly to be compared with data on both NDSS and DENFREE symptomatic cases. In this model, the basic reproduction number, i.e. the number of secondary infections resulting from the introduction of a single infected in an entirely susceptible population, is $R_0^{\mathrm{SEIR}}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$.

$$\begin{aligned}
\frac{\mathrm{dH}_S}{\mathrm{dt}} &= \mu_H N - \beta(t)\frac{(H_I+i)H_S}{N} - \mu_H H_S \\
\frac{\mathrm{dH}_E}{\mathrm{dt}} &= \beta(t)\frac{(H_I+i)H_S}{N} - \sigma H_E - \mu_H H_E \\
\frac{\mathrm{dH}_I}{\mathrm{dt}} &= \sigma H_E - \gamma H_I - \mu_H H_I \\
\frac{\mathrm{dH}_R}{\mathrm{dt}} &= \gamma H_I - \mu_H H_R \\
\frac{\mathrm{dC}_I}{\mathrm{dt}} &= \sigma H_E
\end{aligned} \tag{1}$$

This model is compared with two other models that include the mosquito vector transmission components. In the first one, which is a Ross-Mcdonald type model derived from Pandey et al. (2013), the vector is modelled explicitly with three compartments (Susceptible-Exposed-Infected) (cf. Fig. 2). In the second one, derived from Laneri et al. (2010), the vector is modelled implicitly as an external force of infection including two stages, latent ($\kappa$) and current ($\lambda$) (cf. Fig. 3). We derived $R_0$ for each model as $R_0^{\mathrm{Pandey}}(t) = \frac{\beta_H\beta_V(t)\sigma\tau}{(\gamma+\mu_H)(\sigma+\mu_H)\mu_V(\mu_V+\tau)}$ and the estimation $R_0^{\mathrm{Laneri}}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$ (Champagne et al., 2016). In order to compare these models with the non-vector models, we considered the same definition (i.e. the number of secondary human infections resulting from the introduction of a single infected human in a entirely susceptible population), and not the reproduction ratio per generation provided through the use of the next generation matrix.

The equations describing the Pandey model are:

$$\begin{aligned}
\frac{\mathrm{dH}_S}{\mathrm{dt}} &= \mu_H N - \beta_H v_I H_S - \mu_H H_S \\
\frac{\mathrm{dH}_E}{\mathrm{dt}} &= \beta_H v_I H_S - \sigma H_E - \mu_H H_E \\
\frac{\mathrm{dH}_I}{\mathrm{dt}} &= \sigma H_E - \gamma H_I - \mu_H H_I \\
\frac{\mathrm{dH}_R}{\mathrm{dt}} &= \gamma H_I - \mu_H H_R \\
\frac{\mathrm{dv}_S}{\mathrm{dt}} &= \mu_V - \beta_V(t)\frac{(H_I+i)}{N}v_S - \mu_V v_S \\
\frac{\mathrm{dv}_E}{\mathrm{dt}} &= \beta_V(t)\frac{(H_I+i)}{N}v_S - \tau v_E - \mu_V v_E \\
\frac{\mathrm{dv}_I}{\mathrm{dt}} &= \tau v_E - \mu_V v_I \\
\frac{\mathrm{dC}_I}{\mathrm{dt}} &= \sigma H_E
\end{aligned} \tag{2}$$

where $v_s$ is the proportion of susceptible mosquitoes, $v_E$ the proportion of exposed mosquitoes, and $v_I$ the proportion of infected mosquitoes.

The equations describing the Laneri model are:

$$\begin{aligned}
\frac{\mathrm{dH}_S}{\mathrm{dt}} &= \mu_H N - \lambda H_S - \mu_H H_S \\
\frac{\mathrm{dH}_E}{\mathrm{dt}} &= \lambda H_S - \sigma H_E - \mu_H H_E \\
\frac{\mathrm{dH}_I}{\mathrm{dt}} &= \sigma H_E - \gamma H_I - \mu_H H_I \\
\frac{\mathrm{dH}_R}{\mathrm{dt}} &= \gamma H_I - \mu_H H_R \\
\frac{d\kappa}{\mathrm{dt}} &= \beta(t)\frac{2(H_I+i)\tau}{N} - 2\tau\kappa \\
\frac{d\lambda}{\mathrm{dt}} &= 2\tau\kappa - 2\tau\lambda \\
\frac{\mathrm{dC}_I}{\mathrm{dt}} &= \sigma H_E
\end{aligned} \tag{3}$$

#### 2.2.2. Model with explicit asymptomatic individuals (SEIAR)

We also consider a model in which asymptomatic infections are explicitly taken into account in the transmission process (cf. Fig. 4). In this model, we assume that, after the incubation period, there are three possible manifestations of the disease: asymptomatic ($H_A$), mildly symptomatic not requiring hospitalization ($H_I$) and hospitalized cases ($H_H$). Asymptomatic cases are defined in the dengue study as asymptomatic or pauci-symptomatic (presence of other symptoms not being sufficient to classify as symptomatic). Hospital cases are defined as NDSS cases (reported by the surveillance system in hospitals). We assume that symptomatic DENFREE cases are either ($H_I$) or ($H_H$). We also assume that asymptomatic cases transmit the disease as much as symptomatic cases, as recently shown (Duong et al., 2015), and therefore, $R_0^{\mathrm{SEIAR}}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$. $C_H$, $C_I$, $C_A$ represent respectively the count of new hospitalized, symptomatic and asymptomatic cases, and each is aggregated weekly to be compared respectively with NDSS data, DENFREE data on symptomatic cases, and DENFREE data on asymptomatic cases.

**Table 3**
**Prior distributions of parameters**. "Uniform[0,20]" indicates a uniform distribution in the range [0,20]. "Normal(0.44, 0.05) in [0.2, 1]" indicates a a truncated normal distribution with mean 0.44 and standard deviation 0.05, restricted to the range [0.2,1].

| Parameter | | Prior distribution | Reference | Models |
|---|---|---|---|---|
| *Infectiousness, incubation and mortality rates* | | | | |
| $\gamma^{-1}$ | Infectious period (days) | 4.5 | WHO (2017) | All |
| $\sigma^{-1}$ | Intrinsic incubation period (days) | 5.9 | Chan and Johansson (2012) | Pandey, Laneri |
| $\tau^{-1}$ | Extrinsic incubation period (days) | 10 | Chan and Johansson (2012) | Pandey, Laneri |
| $\sigma^{-1}$ | Both incubation periods (days) | 15.9 | Chan and Johansson (2012) | SEIR, SEIAR, SEIR2, SEIR2psi |
| $\mu_H^{-1}$ | Age duration (years) | 15 | Assumed | All |
| $\mu_V^{-1}$ | Mosquito lifespan (days) | 15 | Liu-Helmersson et al. (2014) | Pandey |
| $\rho_A$ | Proportion of asymptomatic cases | Uniform[0, 1] | Assumed | SEIAR |
| $\rho_H$ | Proportion of hospitalized cases | Uniform[0, 1] | Assumed | SEIAR |
| *Transmission parameters* | | | | |
| $R_0$ | Average basic reproduction number | Uniform[0, 20] | Assumed | All |
| $\beta_V$ | Transmission from human to mosquito | Uniform[0.1, 2] | Pandey et al. (2013) | Pandey |
| $\psi$ | inh./enh. of infectiousness | Uniform[0.5, 3] | Assumed | SEIR2psi |
| *Initial conditions* | | | | |
| $H_I(0)$ | Initial number of infected individuals | Uniform[0, 100] | Assumed | SEIR, Pandey, Laneri |
| $H_E(0)$ | Initial number of exposed individuals | $H_I(0)$ | Assumed | SEIR, Pandey, Laneri, ($3H_I(0)$ in SEIAR) |
| $H_S(0)$ | Initial number of susceptible individuals | N*Normal(0.44, 0.05) in [0.2, 1] | Thai et al. (2005) | All |
| $v_I(0)$ or $\lambda(0)$ | Initial number of infected mosquitoes | 0 | Assumed | Pandey or Laneri |
| $v_E(0)$ or $\kappa(0)$ | Initial number of exposed mosquitoes | 0 | Assumed | Pandey or Laneri |
| $H_H(0)$, $H_A(0)$ | Initial number of asymptomatic and hospitalized individuals | $H_I(0)$ | Assumed | SEIAR |
| $H_{S1}(0)$, $H_{S2}(0)$ | Initial number of individuals susceptibles to 1 strain | N*Uniform[0.01, 0.5] | Assumed | SEIR2, SEIR2psi |
| $H_{I1}(0)$, $H_{I2}(0)$ | Initial number of infected individuals | Uniform[0, 100] | Assumed | SEIR2, SEIR2psi |
| $H_{E1}(0)$, $H_{I21}(0)$, $H_{E21}(0)$ | Initial number of infected and exposed individuals | $H_{I1}(0)$ | Assumed | SEIR2, SEIR2psi |
| $H_{E2}(0)$, $H_{I12}(0)$, $H_{E12}(0)$ | Initial number of infected and exposed individuals | $H_{I2}(0)$ | Assumed | SEIR2, SEIR2psi |
| *Observation process* | | | | |
| $r_N$ | Observation rate for NDSS data | Uniform[0, 1] | Assumed | SEIR, Laneri, Pandey, SEIR2, SEIR2psi |
| $r_H$ | Observation rate for NDSS data | 1 | Assumed | SEIAR |
| $r_D$ | Observation rate for DENFREE data (symptomatic) | Fixed | cf. Table 4 | All |
| $r_A$ | Observation rate for DENFREE data (asymptomatic) | Fixed | cf. Table 4 | SEIAR |
| $\phi$ | Overdispersion | Uniform[0, 1] | Assumed | All |
| *Seasonality parameters* | | | | |
| $b$ | Amplitude of the sinusoidal forcing | Uniform[0, 1] | Assumed | All |
| $p$ | Phase of the sinusoidal forcing | Uniform[−0.5, 0.5] | Assumed | All |
| $i$ | Import parameter | Uniform[0, 10] | Assumed | All |
| *Total population* | | | | |
| $N$ | Total population | 161391 | | All |

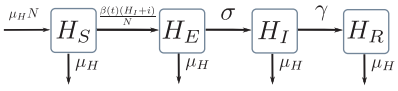


**Fig. 1. Graphical representation of SEIR model.** $H_S$ susceptible individuals; $H_E$ infected (not yet infectious) individuals; $H_I$ infectious individuals; $H_R$ recovered individuals; $\beta(t)$ is the transmission parameter; $\sigma$ is the rate at which $H_E$-individuals move to the infectious class $H_I$; infectious individuals ($H_I$) then recover at rate $\gamma$; individuals leave the children population at rate $\mu_H$. $H_S + H_E + H_I + H_R = N$.

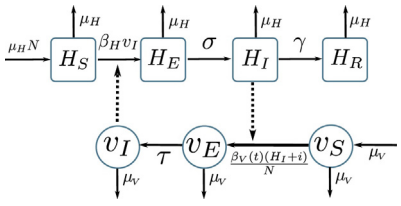


**Fig. 2. Graphical representation of Pandey model (Pandey et al., 2013).** Squared boxes and circles correspond respectively to human and vector compartments. Plain arrows represent transitions from one state to the next. Dashed arrows indicate interactions between humans and vectors. $H_S$ susceptible individuals; $H_E$ infected (not yet infectious) individuals; $H_I$ infectious individuals; $H_R$ recovered individuals; $\beta_H$ is the transmission parameter from vector to human; $\sigma$ is the rate at which $H_E$-individuals move to the infectious class $H_I$; infectious individuals ($H_I$) then recover at rate $\gamma$; individuals leave the children population at rate $\mu_H$; $H_S + H_E + H_I + H_R = N$; $v_S$ proportion of susceptible vectors; $v_E$ proportion of infected (not yet infectious) vectors; $v_I$ proportion of infectious vectors; $\beta_V(t)$ is the transmission parameter from human to vector; $\tau$ is the rate at which $v_E$-vectors move to the infectious class $v_I$; vectors die at rate $\mu_V$.

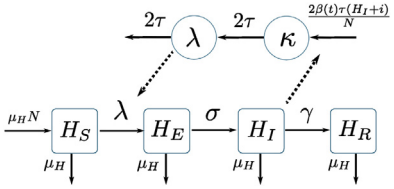


**Fig. 3. Graphical representation of Laneri model (Laneri et al., 2010).** Squared boxes and circles correspond respectively to human and vector compartments. Plain arrows represent transitions from one state to the next. Dashed arrows indicate interactions between humans and vectors. $H_S$ susceptible individuals; $H_E$ infected (not yet infectious) individuals; $H_I$ infectious individuals; $H_R$ recovered individuals; $\sigma$ is the rate at which $H_E$-individuals move to the infectious class $H_I$; infectious individuals ($H_I$) then recover at rate $\gamma$; individuals leave the children population at rate $\mu_H$; $H_S + H_E + H_I + H_R = N$; implicit vector-borne transmission is modelled with the compartments $\kappa$ and $\lambda$; $\lambda$ current force of infection; $\kappa$ latent force of infection reflecting the exposed state for mosquitoes during the extrinsic incubation period; $\beta(t)$ is the transmission parameter; $\tau$ is the transition rate associated with the extrinsic incubation period.

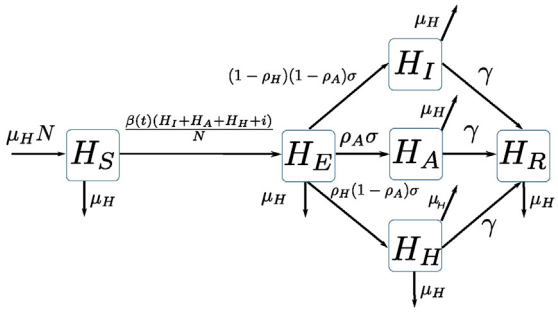


**Fig. 4. Graphical representation of SEIAR model.** $H_S$ susceptible individuals; $H_E$ infected (not yet infectious) individuals; $H_A$ asymptomatic infectious individuals; $H_I$ mildly symptomatic infectious individuals; $H_H$ hospitalized infectious individuals; $H_R$ recovered individuals; $\beta(t)$ is the transmission parameter; $\sigma$ is the rate at which $H_E$-individuals move to the infectious classes $H_I$, $H_A$ and $H_H$; a proportion $\rho_A$ of $H_E$-individuals do not show symptoms during the infectious period; a proportion $\rho_H$ of symptomatic individuals go to hospital; infectious individuals ($H_I$,$H_A$,$H_H$) then recover at rate $\gamma$; individuals leave the children population at rate $\mu_H$. $H_S + H_E + H_A + H_I + H_H + H_R = N$.

$$
\begin{aligned}
\frac{dH_S}{dt} &= \mu_H N - \beta(t)\frac{(H_I + H_A + H_H + i)H_S}{N} - \mu_H H_S\\
\frac{dH_E}{dt} &= \beta(t)\frac{(H_I + H_A + H_H + i)H_S}{N} - \sigma H_E - \mu_H H_E\\
\frac{dH_A}{dt} &= \rho_A \sigma H_E - \gamma H_A - \mu_H H_A\\
\frac{dH_H}{dt} &= \rho_H(1-\rho_A)\sigma H_E - \gamma H_H - \mu_H H_H\\
\frac{dH_I}{dt} &= (1-\rho_H)(1-\rho_A)\sigma H_E - \gamma H_I - \mu_H H_I\\
\frac{dH_R}{dt} &= \gamma(H_I + H_A + H_H) - \mu_H H_R\\
\frac{dC_I}{dt} &= (1-\rho_A)\sigma H_E\\
\frac{dC_A}{dt} &= \rho_A \sigma H_E\\
\frac{dC_H}{dt} &= \rho_H(1-\rho_A)\sigma H_E
\end{aligned}
\tag{4}
$$

### 2.2.3. Model with two virus serotypes

In the 2012 and 2013 epidemics, DENV-1 was highly dominant: the three other serotypes represented about 2% of the cases reported in the DENFREE study in 2012 and about 21% in 2013 (cf. Table 1). Therefore, a two-strain model is also studied, in which we separate DENV-1 cases from DENV-2, DENV-3 and DENV-4 combined (cf. Fig. 5). For simplicity and parsimony in the number of parameters, the two strains share the same parameter values. We first assume both strains to be independent ($\psi = 1$ in Eq. (5), called SEIR2 model). In this context, the reproduction numbers for each strain are equal, $R_0^{\mathrm{SEIR2}_1}(t) = R_0^{\mathrm{SEIR2}_2}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$.

$$
\begin{aligned}
\frac{dH_S}{dt} &= \mu_H N - \beta(t)\frac{(H_{I1} + \psi H_{I21} + i)H_S}{N} - \beta(t)\frac{(H_{I2} + \psi H_{I12} + i)H_S}{N} - \mu_H H_S\\
\frac{dH_{E1}}{dt} &= \beta(t)\frac{(H_{I1} + \psi H_{I21} + i)H_S}{N} - \sigma H_{E1} - \mu_H H_{E1}\\
\frac{dH_{I1}}{dt} &= \sigma H_{E1} - \gamma H_{I1} - \mu_H H_{I1}\\
\frac{dH_{S1}}{dt} &= \gamma H_{I1} - \beta(t)\frac{(H_{I2} + \psi H_{I12} + i)H_{S1}}{N} - \mu_H H_{S1}\\
\frac{dH_{E12}}{dt} &= \beta(t)\frac{(H_{I2} + \psi H_{I12} + i)H_{S1}}{N} - \sigma H_{E12} - \mu_H H_{E12}\\
\frac{dH_{I12}}{dt} &= \sigma H_{E12} - \gamma H_{I12} - \mu_H H_{I12}\\
\frac{dH_{E2}}{dt} &= \beta(t)\frac{(H_{I2} + \psi H_{I12} + i)H_S}{N} - \sigma H_{E2} - \mu_H H_{E2}\\
\frac{dH_{I2}}{dt} &= \sigma H_{E2} - \gamma H_{I2} - \mu_H H_{I2}\\
\frac{dH_{S2}}{dt} &= \gamma H_{I2} - \beta(t)\frac{(H_{I1} + \psi H_{I21} + i)H_{S2}}{N} - \mu_H H_{S2}\\
\frac{dH_{E21}}{dt} &= \beta(t)\frac{(H_{I1} + \psi H_{I21} + i)H_{S2}}{N} - \sigma H_{E21} - \mu_H H_{E21}\\
\frac{dH_{I21}}{dt} &= \sigma H_{E21} - \gamma H_{I21} - \mu_H H_{I21}\\
\frac{dH_R}{dt} &= \gamma(H_{I12} + H_{I21}) - \mu_H H_R\\
\frac{dC_I}{dt} &= \sigma(H_{E1} + H_{E2} + H_{E21} + H_{E12})\\
\frac{dC_{I1}}{dt} &= \sigma(H_{E1} + H_{E21})\\
\frac{dC_{I2}}{dt} &= \sigma(H_{E2} + H_{E12})
\end{aligned}
\tag{5}
$$

We also considered another version of the model including interaction between strains (ten Bosch et al., 2016), in order to reflect the fact that secondary infection with a heterologous serotype leads more often than primary infection to severe manifestations of the disease (Halstead, 2007). In our model (called SEIR2psi model), primary and secondary infections differ in infectiousness, through a parameter $\psi$ (Ferguson et al., 1999). This parameter is estimated between 0.5 and 3: values superior to 1 correspond to transmission cross-enhancement (because of higher virus titers during secondary infections (Ferguson et al., 1999)) and values inferior to 1 suggest a lower infectivity for secondary infected individuals (for example because they are hospitalized and less in contact with the population (Aguiar et al., 2011)). As in Ferguson et al. (1999), we define $R_0^{\mathrm{SEIR2}_1}(t) = R_0^{\mathrm{SEIR2}_2}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$ the basic reproduction number for each strain.

$C_I$ represents the count of new cases for both serotypes and it is aggregated weekly to be compared with NDSS data. $C_{I1}$ and $C_{I2}$ represent the count of new cases for strain 1 and strain 2, and are aggregated weekly to be compared with symptomatic cases from the DENFREE study for DENV-1 and DENV-2/ DENV-3/ DENV-4 respectively.

### 2.2.4. Seasonality

All models include seasonality through the use of a time-varying transmission parameter $\beta(t) = \beta\left[1 + b.\sin\left(2\pi\left(\frac{t}{365} + p\right)\right)\right]$, according to a sinusoidal function whose phase $p$ and amplitude $b$ are estimated.

We also assume that a constant number of cases $i$ are imported.

## 2.3. Prior distributions

The prior distributions are listed in Table 3.

Dirac priors based on the literature were used for the durations of infectiousness and incubation, as well as the mortality rates. In models without vectorial transmission, the incubation period is assumed to be the sum of the extrinsic (in mosquito) and intrinsic (in human) incubation periods, to reflect the generation time of the disease. For transmission parameters, we used wide weakly informative priors.

### 2.3.1. Initial conditions

The initial number of infected individuals is assumed to be equal to

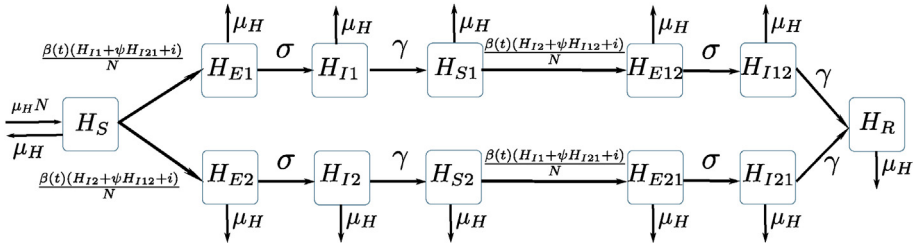


**Fig. 5. Graphical representation of SEIR2 models.** $H_S$ individuals susceptible to both strains; $H_{E1}$ (resp. $H_{E2}$) individuals infected (not yet infectious) to strain 1 (resp. strain 2); $H_{I1}$ (resp. $H_{I2}$) individuals infectious to strain 1 (resp. strain 2); $H_{S1}$ (resp. $H_{S2}$) individuals immune to strain 1 only (resp. strain 2); $H_{E12}$ (resp. $H_{E21}$) individuals (not yet infectious) with a secondary infection to strain 2 (resp. strain 1); $H_{I12}$ (resp. $H_{I21}$) infectious individuals with a secondary infection to strain 2 (resp. strain 1); $H_R$ individuals immune to both strains; $\beta(t)$ is the transmission parameter; $\sigma$ is the rate at which exposed individuals move to the infectious class; infectious individuals then recover at rate $\gamma$; $\psi$ is the change in infectivity for secondary infected individuals in SEIR2psi model (in SEIR2 model, $\psi = 1$); individuals leave the children population at rate $\mu_H$. $H_S + H_{E1} + H_{E2} + H_{I1} + H_{I2} + H_{S1} + H_{S2} + H_{E12} + H_{E21} + H_{I12} + H_{I21} + H_R = N$.

the number of exposed individuals and to be lower than 100, as the model starts in January, during the epidemic trough. Except for the initial proportion of susceptibles, all priors on initial conditions are uniform distributions.

The initial proportion of susceptibles is an influential parameter on the model outputs. It is highly correlated to the transmission parameter $\beta$ (and therefore to the basic reproductive number), which makes it difficult to estimate them both. An informative gaussian prior was therefore used on $H_S(0)$. To date, no large scale seroprevalence study is available for Cambodia, and we relied on a study conducted among schoolchildren in rural Vietnam (Thai et al., 2005), which we considered as the closest setting to be compared with Kampong Cham. We extrapolated their results on schoolchildren (7 to 14 years old) to a 1–15 years old population as follows, where $S_\lambda$ is the proportion of susceptibles among 1–15 year-old children (using their estimation $\lambda = 0.117$):

$$S_\lambda = \frac{1}{15}\sum_{a=1}^{15} \exp(-\lambda a) = 0.44$$

$S_\lambda$ is used as the mean of the gaussian prior, and the standard deviation is fixed at 0.05.

### 2.4. Estimation

#### 2.4.1. Observation model

The propagation models are related to the observed data using a negative binomial observation model (Bretó et al., 2009). The observed number of cases during week $k$ in each dataset, $C_{\text{obs}}^{(k)}$, is assumed to be drawn from a a negative binomial distribution with mean $rC^{(k)}$ and variance $rC^{(k)} + (rC^{(k)})^2\phi$, where $C^{(k)}$ is the total number of new cases simulated by the model during week $k$, $r$ is the observation rate quantifying the amount of non reported cases and $\phi$ is an overdispersion parameter. The quantity $C$ and observation rate $r$ associated to each dataset are indicated in Table 1.

For each epidemiological model, estimations are performed simultaneously on NDSS and DENFREE data (DENFREE data being decomposed into two time series in SEIAR, SEIR2 and SEIR2psi models). The observations of all time series are assumed independent conditional on the underlying disease process, and the model likelihood is the product of the likelihoods calculated on each series.

The observation rate is estimated for NDSS data but fixed for DENFREE data, using informations from the sampling scheme (cf. Table 4), in order to account for the difference in coverage during outbreak investigation between 2012 and 2013. We assume that index cases are all reported and that the observation rate for community cases equals the ratio of people tested over the population of the area. We then extrapolate this observation rate to the total population of the four districts.

Different observation rates are used in the SEIAR model. As NDSS data are interpreted as hospitalized cases, the observation rate is fixed to 1 (which assumes that all hospitalized cases go through surveillance and neglects the presence of private hospitals or non reports from hospitals), but the proportion of hospitalized cases is estimated. For DENFREE cases, the observation rate for symptomatic and asymptomatic individuals is fixed as indicated in Table 4.

The overdispersion parameter $\phi$ is estimated along with the other parameters.

#### 2.4.2. Estimation with Markov Chain Monte Carlo

Models are considered in deterministic framework and estimations are made using random walk Metropolis Hastings. SSM software (Dureau et al., 2013) is used for simulations and calculations. The covariance matrix of the proposal distribution was initialized using adaptative MCMC as in Dureau et al. (2013).

Due to unknown initial conditions, the posterior distribution is multimodal, especially in two strain models. This implies that the adaptive MCMC remains trapped in local maxima and global exploration of the parameter space was performed with latin hypercube sampling. We ran a simplex algorithm on 10,000 parameter sets sampled with latin hypercube sampling (with *lhs* R package (Carnell, n.d.)) and chose the one with the highest posterior value as initialization of the adaptive MCMC algorithm. The results associated to this best fitting equilibrium are displayed in the main text, but other examples of good fitting equilibria for two-strain models are indicated in Appendix E.

### 2.5. Model comparison

#### 2.5.1. Statistical indicators

In order to identify the best model, the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) is used. DIC is an indicator that combines a measure of model fit and a penalty on model complexity, commonly used with MCMC estimations. The best model is the one with the smallest DIC. As it does not enable comparison of models with differing number of observations, we also calculate the RMSE (root mean square error) between observations $C_{\text{obs}}^{(k)}$ and simulations of the model with parameters sampled in the posterior distribution $C^{(k,i)}$, for weeks $k = k_0, \ldots, K$ and simulation $i$:

$$\text{RMSE}_i = \sqrt{\frac{1}{K-k_0}\sum_{k=k_0}^{K}(C^{(k,i)} - C_{obs}^{(k)})^2}$$

We then compute the mean and the 2.5% and 97.5% quantiles of this indicator among the simulated trajectories. We first calculated them on the data used for estimation, separating NDSS data and DENFREE data. For DENFREE data we considered the total number of symptomatic cases to enable comparison of all models: therefore, for this metric, asymptomatic DENFREE cases were not taken into account in SEIAR model, and cases of all serotypes were aggregated in two-strain models. Then we used these indicators to assess the predictive performance of the models, comparing projections of the model with NDSS observations for 2014 and 2015.

**Table 4**
Calculation of the observation rate for DENFREE data.

| Children < 15 years old | 2012 | | 2013 | |
|---|---|---|---|---|
| | Symptomatic | Asymptomatic | Symptomatic | Asymptomatic |
| Population in investigated villages ($n$) | 65,208 | | 65,208 | |
| Population in the 4 districts ($N$) | 161,391 | | 161,391 | |
| Index cases ($a$) | 151 | 0 | 376 | 0 |
| Observation rate for index cases | 1 | . | 1 | . |
| Children tested in communities ($b$) | 1722 | | 4119 | |
| DENV positive in communities ($c$) | 85 | 5 | 198 | 28 |
| Observation rate for community cases ($d = b/n$) | 0.0264 | | 0.0632 | |
| Extrapolated number of community cases ($e = c/d$) | 3219 | 189 | 3135 | 443 |
| Extrapolated total number of cases ($f = a + e$) | 3370 | 189 | 3511 | 443 |
| Observed total number of cases ($g = a + c$) | 236 | 5 | 574 | 28 |
| **Observation rate** ($r_D = (g/f) * (n/N)$ and $r_A = b/N$) | **0.0283** | **0.0107** | **0.0661** | **0.0255** |

*2.5.2. Epidemiological indicators*

Models are also compared according to several indicators to describe their epidemiological behaviour. The basic reproduction number, the observation rate and the initial proportion of susceptibles are estimated using the MCMC chain. In the model with asymptomatic infections we report the estimated proportions of asymptomatic and hospitalized cases.

With parameters sampled in the posterior distribution, we can also re-simulate the model to study hidden states, such as the susceptible and infected classes. The effective reproduction number ($R_e$) is calculated as the seasonal basic reproduction number multiplied by the proportion of susceptibles at each time step, as indicated in Table 5. We then calculate the annual incidence proportion as the total number of infections over one year divided by the total population of susceptibles at the beginning of the year. In models with two strains, we separate the annual incidence of primary infections (as the total number of primary infections over one year divided by the total population of naive individuals at the beginning of the year) and secondary infections (as the total number of secondary infections over one year divided by the total population of susceptibles to one strain only at the beginning of the year).

Calculations are made using R version 3.2.2 (Team, 2015), and graphics using *ggplot2* (Wickham, 2009).

## 3. Results

The calculations are based on the MCMC chain associated with the highest posterior after LHS exploration. In multistrain models, the posterior distribution is multimodal, and examples of other local posterior modes are displayed in Appendix E.

### 3.1. Statistical comparison

For single strain models, the SEIR model proved the best with the DIC criterion (cf. Table 6). For two strain models, the SEIR2psi proved best. As regards simulation-based indicators on the 2002–2014 data, SEIR, Laneri, Pandey models have RMSE values in the same order of magnitude. Indeed, they produce a similar dynamic with respect to the 2002–2014 data (cf. Figs. 6–7), with a period of approximately six years, and a large overestimation of the 2002 outbreak (the mean RMSE for 2002 is far higher ($> 40$) than the average for the other years ($< 25$)). Due to the small number of observed asymptomatic cases, SEIAR model also produces a similar dynamic and the SEIR2 model, despite a larger number of parameters, does not improve over single-strain models. The model with two interacting strains (SEIR2psi) outperforms the other models in terms of RMSE, but the difference is mainly explained by the first year of simulation (2002). However, the RMSE indicator does not penalize for the complexity of this more flexible model, and the differences are not very pronounced given the variability in RMSE values. When visualizing the simulations compared to the data (cf. Figs. 6–7), all models underestimate a large number of epidemic peaks and all models but the one with two interacting strains overpredict the first epidemic peak (as pointed out in Table 6). SEIR2psi is the model in which the large epidemic in 2007 is best reproduced and overall, SEIR2psi is the model that reproduces most accurately the observed data. Moreover, the estimated overdispersion is lower in this model (cf. Table 7), indicating that less observation noise is required to explain the data and that the model explains a larger part of the variability in the data.

Considering the predictive capacity of the models, all models overestimate the 2014 small epidemic, and to a lesser extent the one in 2015 (cf. Fig. 8). In the SEIR2psi model, this overestimation is less pronounced, and the associated RMSE is also smaller (cf. Table 6).

### 3.2. Epidemiological comparison

The average $R_0$ is estimated to be between 2 and 3 in most of the models and the maximum value between 3 and 4 (cf. Table 7), except with the Pandey model, in which it is higher (mean value above 3 and maximum value above 6). The estimated values are very close in the SEIR, Laneri and SEIAR models. These values relate to transmission among children only and we expect therefore the $R_0$ in the whole population to be higher, since children represent approximately one third of the total population, but more than 90% of the total cases observed. Estimates for $R_0$ in South East Asia in the literature range from 1 to 5 (Imai et al., 2015, 2016), with sometimes values above 10 (Johansson et al., 2011), displaying large variations in time and space, and in particular, the estimates for $R_0$ in Cambodia based on age-stratified

**Table 5**
**Reproduction numbers calculation in each model**. Calculations are based on the next generation matrix method (Diekmann et al., 2010).

| | SEIR, Laneri, SEIAR | Pandey | SEIR2 | SEIR2psi |
|---|---|---|---|---|
| $R_0(t)$ | $\frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$ | $\frac{\beta_H\beta_V(t)\sigma\tau}{(\gamma+\mu_H)(\sigma+\mu_H)\mu_V(\mu_V+\tau)}$ | $\frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$ | $\frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$ |
| $R_e(t)$ (or $R_e^i$, $i = 1, 2$) | $R_0(t)\frac{H_S(t)}{N}$ | $R_0(t)\frac{H_S(t)}{N}v_s(t)$ | $R_0(t)\frac{H_S(t)+H_{Sj}(t)}{N}$ | $R_0(t)\frac{H_S(t)+\psi H_{Sj}(t)}{N}$ |

**Table 6**
**Information criteria in deterministic models.** DIC is the Deviance Information Criterion (Spiegelhalter et al., 2002). RMSE is the root mean square error (mean and 2.5% and 97.5% quantiles) between simulations and observations: it is calculated separately on the datasets used for estimations (NDSS data for 2002–2013 and DENFREE data for 2012–2013) and on the test set (NDSS data for 2014–2015). It is also computed on separated years in order to highlight well or badly estimated years: for example, for each simulation $i$, $\frac{1}{12}(\text{RMSE}_i^{\text{NDSS 2002}})^2 + \frac{11}{12}(\text{RMSE}_i^{\text{NDSS 2003-2013}})^2 = (\text{RMSE}_i^{\text{NDSS}})^2$. Calculations are based on the MCMC chain associated with the highest posterior.

| Model | SEIR | Laneri | Pandey | SEIAR | SEIR2 | SEIR2psi |
|---|---|---|---|---|---|---|
| nb parameters | 8 | 8 | 9 | 9 | 11 | 12 |
| nb observations | 665 | 665 | 665 | 704 | 704 | 704 |
| *ESTIMATION SET* | | | | | | |
| DIC | 4343 | 4360 | 4364 | 4472 | 4492 | 4446 |
| RMSE NDSS | 25 (22–28) | 25 (22–29) | 26 (23–29) | 26 (23–29) | 25 (22–29) | 21 (19–24) |
| RMSE NDSS 2002 | 43 (28–59) | 42 (27–58) | 45 (29–63) | 43 (27–61) | 40 (24–61) | 17 (12–25) |
| RMSE NDSS 2003–2013 | 23 (20–25) | 23 (21–26) | 23 (21–25) | 23 (21–26) | 24 (21–26) | 22 (19–24) |
| RMSE DENFREE | 24 (19–29) | 24 (19–29) | 24 (19–29) | 24 (19–29) | 23 (19–27) | 25 (19–31) |
| *TEST SET* | | | | | | |
| RMSE 2014–2015 | 15 (11–19) | 15 (11–19) | 16 (12–20) | 15 (11–20) | 17 (13–22) | 13 (10–16) |
| RMSE 2014 | 16 (11–23) | 16 (11–23) | 17 (11–24) | 17 (11–24) | 17 (11–24) | 14 (10–19) |
| RMSE 2015 | 13 (9–17) | 13 (9–17) | 13 (10–18) | 13 (10–17) | 17 (12–23) | 12 (9–16) |

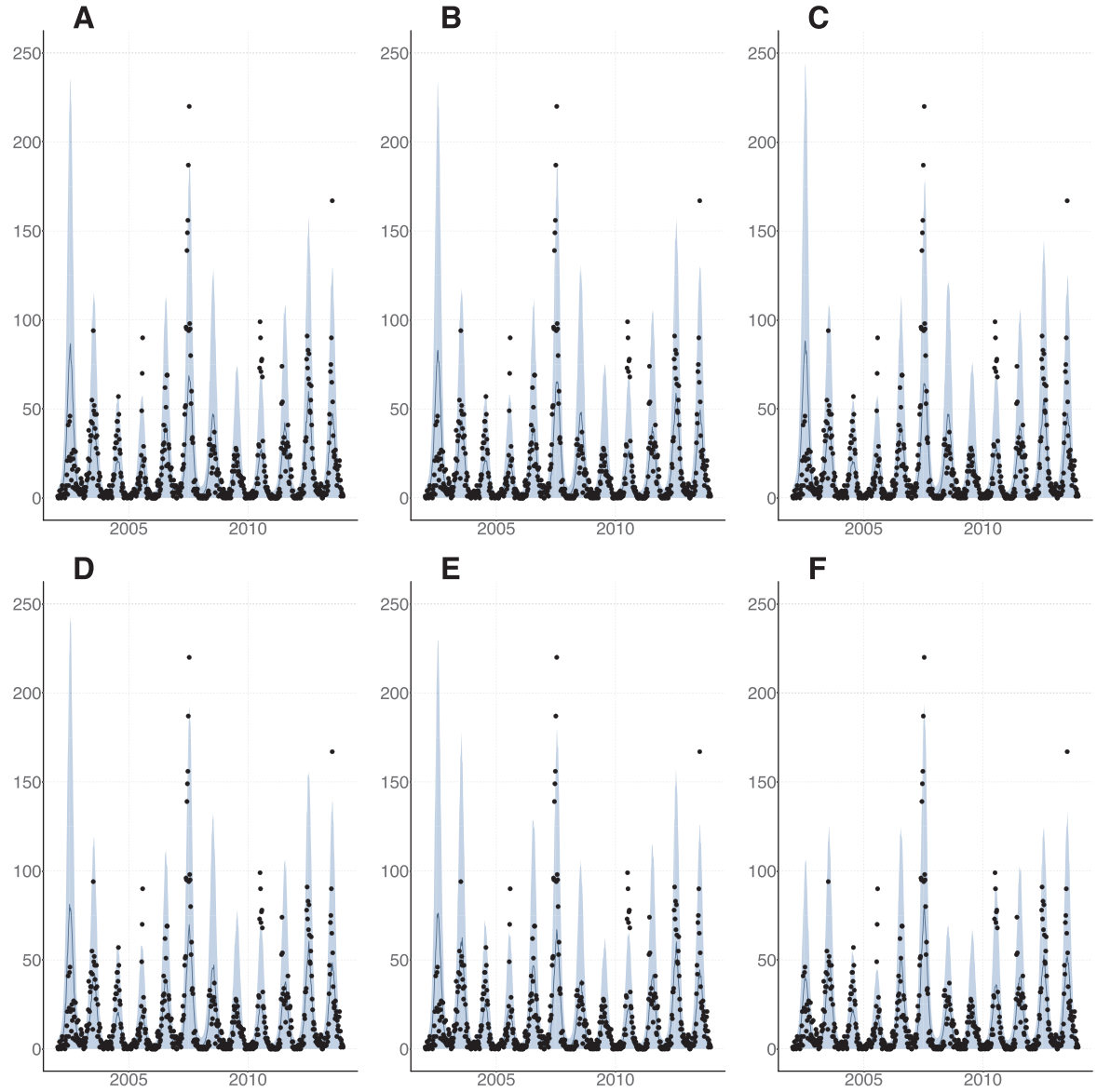


**Fig. 6. Number of observed cases per week and NDSS data, 2002–2013.** Simulations with negative binomial noise using parameters from the MCMC chain associated with the highest posterior, calculated using both NDSS and DENFREE datasets. Posterior median (solid line), 95% credible intervals (shaded blue area) and NDSS data points (black dots). A. SEIR model. B. Laneri model. C. Pandey model. D. SEIAR model. E. SEIR2 model. F. SEIR2psi model.

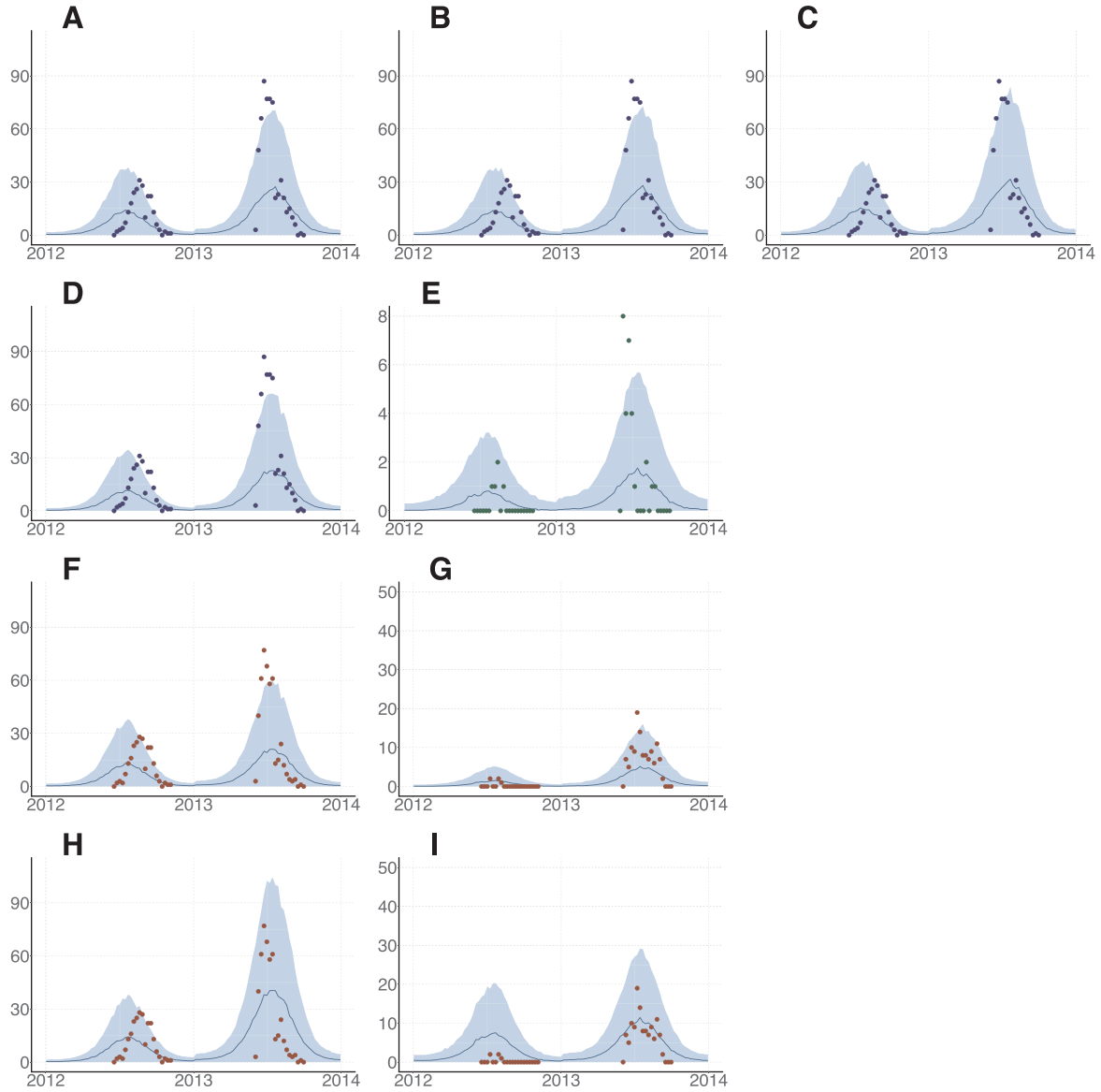


**Fig. 7. Number of observed cases per week and DENFREE data, 2012–2013.** Simulations with negative binomial noise using parameters from the MCMC chain associated with the highest posterior, calculated using both NDSS and DENFREE datasets. Posterior median (solid line), 95% credible intervals (shaded blue area) and DENFREE data points (blue dots: symptomatic cases, green dots: asymptomatic cases, red dots: serotype specific symptomatic cases). A. SEIR model. B. Laneri model. C. Pandey model. D. SEIAR model, symptomatic cases. E. SEIAR model, asymptomatic cases. F. SEIR2 model, cases due to strain 1 and DENV1 observations. G. SEIR2 model, cases due to strain 2 and DENV2/DENV3/DENV4 observations. H. SEIR2 model with interactions, cases due to strain 1 and DENV1 observations. I. SEIR2 model with interactions, cases due to strain 2 and DENV2/DENV3/DENV4 observations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

case-notification data are between 2 and 7 (Imai et al., 2016), so that our estimates are in the same order of magnitude. Nevertheless, our estimation of $R_0$ strongly depends on the estimation of the initial proportion of susceptibles ($H_S(0)$), which is unknown in the case of Cambodia and can bias the estimates. It is indeed very difficult to evaluate $R_0$ via incidence data (Imai et al., 2016), especially for endemic diseases without information on seroprevalence, and it may be more informative to study the effective reproductive number ($R_e$), which accounts for the proportion of immune individuals in the population. The estimated $R_e$ has a mean value around 1 and a maximum value around 1.7 in most models, but it is higher in the Pandey model (the time-varying behaviour of $R_e$ is displayed in Appendix D).

In SEIR2psi model, the parameter $\psi$ quantifying the interaction between strains is inferior to 1, suggesting a reduced infectivity of secondary infections on average, as in Aguiar et al. (2011), or in Coudeville and Garnett (2012). This suggests that, as far as infectiousness is concerned, cross-protection among serotypes is more important than cross-enhancement to explain the results observed in the field.

It is estimated in the models that approximately half of the children are susceptible to the disease, which is close to the informative prior used, highlighting the difficulty to estimate this quantity due to identifiability issues. This proportion is lower in the Pandey model. In the models with two strains, the number of susceptibles differs between models, but as a whole, at least 50% of children are not totally immune to the disease and can still suffer infection. The proportion of children who are susceptible to one or both strains are however correlated in the MCMC chain (cf. Appendix), indicating that their relative proportions are not well identified. These values are in the range of the measures of seroprevalence in several Asian countries (L'Azou et al., 2016, 2018). As the measures reveal large differences between countries (L'Azou et al., 2016, 2018; Prayitno et al., 2017; Thai et al., 2005;

**Table 7**
**Epidemiological criteria among children under 15 years old.** Estimated parameters and indicators based on simulations over 2002–2015 (posterior median and 95% credible intervals). Calculations are based on the MCMC chain associated with the highest posterior.

| Model | | SEIR | Laneri | Pandey | SEIAR | SEIR2 | SEIR2 psi |
|---|---|---|---|---|---|---|---|
| Mean $R_0$ | Median (95%CI) | 2.32 (2.25–2.41) | 2.38 (2.3–2.46) | 3.32 (3.16–3.47) | 2.3 (2.23–2.37) | 1.45 (1.42–1.48) | 2.47 (2.35–2.57) |
| Max $R_0$ | Median (95%CI) | 3.64 (3.51–3.77) | 3.66 (3.53–3.8) | 5.87 (5.57–6.17) | 3.6 (3.48–3.72) | 2.29 (2.23–2.35) | 3.78 (3.59–3.94) |
| $\psi$ | Median (95%CI) | | | | | | 0.67 (0.62–0.73) |
| $H_S(0)/N$ (%) | Median (95%CI) | 46 (45–47) | 46 (45–47) | 37 (36–38) | 47 (46–47) | 49 (45–56) | 35 (30–38) |
| $H_{S1}(0)/N$ (%) | Median (95%CI) | | | | | 22 (14–25) | 14 (10–21) |
| $H_{S2}(0)/N$ (%) | Median (95%CI) | | | | | 27 (19–30) | 4 (1–11) |
| Observation rate (%) | Median (95%CI) | 12 (11–13) | 12 (11–13) | 10 (9–11) | | 11 (10–12) | 7 (6–7) |
| Hospitalized (%) | Median (95%CI) | | | | 14 (13–15) | | |
| Asymptomatic (%) | Median (95%CI) | | | | 15 (13–19) | | |
| Over-dispersion | Median (95%CI) | 0.73 (0.59–0.91) | 0.75 (0.61–0.93) | 0.7 (0.57–0.87) | 0.9 (0.74–0.99) | 0.73 (0.59–0.91) | 0.56 (0.45–0.76) |
| Median annual incidence proportion | | | | | | | |
| Primary infection (%) | Median 2002–2015 (min–max) | 8 (5–16) | 8 (4–16) | 12 (7–25) | 8 (4–16) | 6 (3–10) | 14 (7–27) |
| Secondary infection (%) | Median 2002–2015 (min–max) | | | | | 3 (1–5) | 7 (3–14) |
| Mean $R_e$ | Median (95%CI) | 1.03 (1.03–1.03) | 1.06 (1.06–1.07) | 1.14 (1.13–1.15) | 1.03 (1.03–1.03) | | |
| Mean $R_e$ strain 1 | Median (95%CI) | | | | | 1.03 (1.02–1.03) | 1.03 (1.02–1.03) |
| Mean $R_e$ strain 2 | Median (95%CI) | | | | | 1.04 (1.03–1.04) | 1.02 (1.02–1.03) |
| Max $R_e$ | Median (95%CI) | 1.7 (1.66–1.74) | 1.72 (1.69–1.77) | 2.18 (2.12–2.27) | 1.7 (1.67–1.74) | | |
| Max $R_e$ strain 1 | Median (95%CI) | | | | | 1.73 (1.69–1.77) | 1.74 (1.71–1.77) |
| Max $R_e$ strain 2 | Median (95%CI) | | | | | 1.72 (1.69–1.75) | 1.69 (1.65–1.73) |

Vongpunsawad et al., 2017), a seroprevalence survey in Cambodia would be particularly useful to evaluate which scenario is more plausible.

The observation rate for NDSS data varies between models, from 6 to 13%. These values indicate that a large proportion of dengue cases are not reported in national surveillance, likely reflecting mild symptoms that do not require hospitalization or misdiagnosis or misreporting.

In the SEIR2psi model, we also plotted the current number of infected individuals for each strain (cf. Fig. 9). In our simulations, the first strain is responsible for large explosive outbreaks, whereas the second one has a more regular pattern over the years. Moreover, the two strains are asynchronous and each one dominates for two or three years. It is also qualitatively close to the dynamics observed in Thailand (Reich et al., 2013) or Singapore (Ong et al., 2017).

The proportion of susceptible individuals also displays asynchronous dynamics between strains, as they reflect the history of past epidemics. Despite the seasonality and the year-to-year variations, the total number of susceptibles remains high (cf. Fig. 9 for SEIR2psi and SEIR models), allowing the possibility for large outbreaks to occur in the future. However, multi-strain models are difficult to parameterize due to the uncertainty in the initial conditions, and models with differing trajectories and parameters may have close performances in terms of model fitting (cf. Appendix E).

We calculated the annual incidence proportion as the proportion of new infections over one year among the susceptibles at the beginning of that year (cf. Fig. 10). The values for primary infections are coherent with other studies in Vietnam or Indonesia who analyzed seroprevalence data or seroconversion data (Graham et al., 1999; Prayitno et al., 2017; Thai et al., 2005, 2007; Tien et al., 2010). The incidence proportion is highly variable from one year to the next, especially in models with two interacting strains.

## 4. Discussion

With two datasets reporting dengue cases in the Kampong Cham region in Cambodia, we compared several models to represent dengue transmission dynamics in a rural setting. In order to assess the quality of the models, we compared their statistical properties and analysed their epidemiological features.

The best model describing the dengue trend over 14 years of data was the two strain model, with reduced infectivity for secondary infected individuals. Secondary infections being more prone to severe dengue (Halstead, 2007), these individuals may stay at home or at the hospital, and be less involved in the spread of the disease than the ones whose mild symptoms do not alter what Stoddard et al. (2013) call "house-to-house movement" (Yoon et al., 2012; Perkins et al., 2016). There is also some indication from studies on the infectiousness of infected people to mosquitoes, that symptomatic infections are less infectious than asymptomatic individuals (Duong et al., 2015). Likewise there was a small negative impact of IgG on transmission rates that would be indicative of secondary and potentially more symptomatic cases (Nguyen et al., 2013). This feature was previously analytically studied and coherent with dengue incidence time series in Thailand (Aguiar et al., 2011, 2014).

On the contrary, including vectorial transmission or a compartment for asymptomatic infections did not seem to improve the model fit despite the additional complexity. In our models, the role of mosquitoes in transmission seems well captured by the seasonal forcing implemented in the model. The limited added-value of including vectorial components has been observed by several authors previously, when studying one dengue season only (Pandey et al., 2013) or when a seasonal forcing was included in the model (Rocha et al., 2016). Mathematical analyses have suggested that because the time scale of the mosquito epidemiology is so fast compared to that in humans, it will be slaved by the slower human epidemiology (Rocha et al., 2013). Thus, for understanding human disease epidemiology, mainly the dynamics of the human time scale are essential and inclusion of mosquito dynamics results in an unnecessary increase in model complexity when vector data is not available (Dye and Williams, 1995; Rocha et al., 2013, 2016), and when potential seasonal variations are already accounted for.

Our models account for underreporting, and the lack of additional improvement when including explicitly the asymptomatic class is likely due to the very few asymptomatic infections observed, which may be due to a very strict definition of asymptomatic infections in the DENFREE protocol. Yoon et al. (2012) also observe many inapparent cases in their cohort but few strictly asymptomatic cases in their cluster

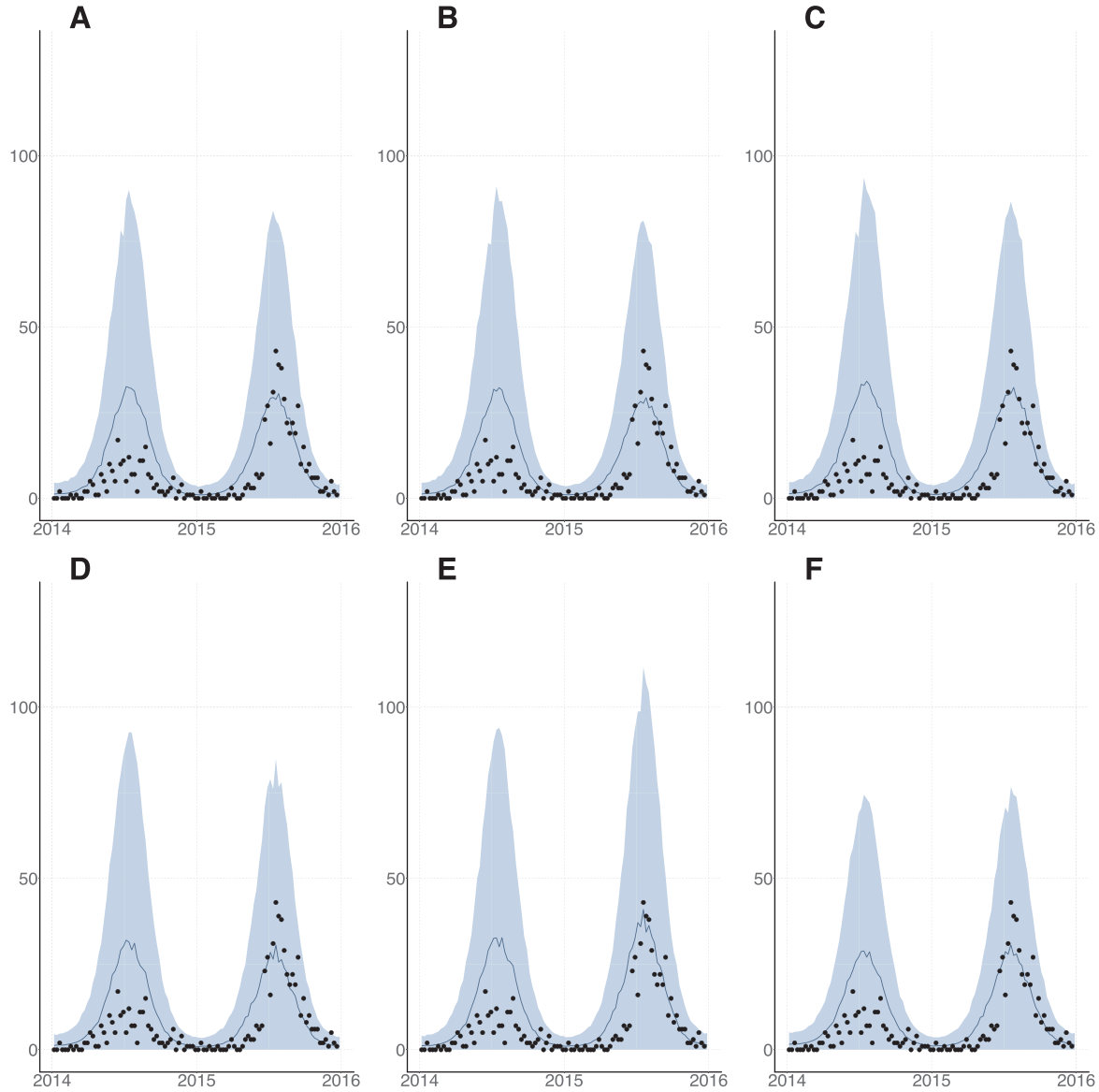


**Fig. 8. Projections of the number of observed cases per week and NDSS data, 2014–2015.** Simulations with negative binomial noise using parameters from the MCMC chain associated with the highest posterior, calculated using both NDSS and DENFREE datasets. Posterior median (solid line), 95% credible intervals (shaded blue area) and NDSS data points (black dots). A. SEIR model. B. Laneri model. C. Pandey model. D. SEIAR model. E. SEIR2 model. F. SEIR2psi model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

study in Thailand. Therefore, in our model, including a compartment for asymptomatic individuals had only little influence on the overall transmission due to their small number.

However, even though the two strain model with interactions reproduces the observed data more adequately, it is difficult to parameterize. Differing model parameterizations may lead to differing dynamics that relate closely to the observed data, as shown in Appendix E, indicating identifiability issues (Miao et al., 2011). Identifiability is nonetheless a central problem in designing and estimating epidemiological models, that also affects models with vector-borne transmission (Kao and Eisenberg, 2018) and even simple models (Evans et al., 2002). In our case, this problem was related to two phenomena. On the one hand, some parameters values were correlated, and therefore difficult to distinguish from one another. On the other hand, in some models, the posterior distribution was multimodal, and therefore both hard to explore with many inference methods and hard to interpret given the presence of several possible trajectories. In order to limit these difficulties in all models, we fixed some parameters (in particular the durations of infection and incubation, or the observation rate in DENFREE data), and relied on informative priors (in particular for the initial proportion of susceptibles). In addition, it is possible to rule out some trajectories that correctly fit the data but display epidemiological incoherences in the qualitative dynamics of their hidden states (such as the depletion of the susceptible population or the absence of serotype asynchrony), in the spirit of pattern matching methods (ten Bosch et al., 2016). We therefore studied the epidemiological characteristics of the selected models to ensure such coherence. Other trajectories may, nevertheless, remain equally plausible given the available information and require additional data to be further explored. These issues are even more important when evaluating public health measures, as models with similar fit may have different responses to control scenarios (Kao and Eisenberg, 2018; ten Bosch et al., 2016). Therefore, care is required when choosing the best model, and the uncertainty in the model structure must also be taken into account, especially when the observed data are scarce.

Over the six models that were explored, we also obtained some insight on the parameters describing transmission. The average annual $R_0$ in our paediatric population is estimated between 1 and 4 over all

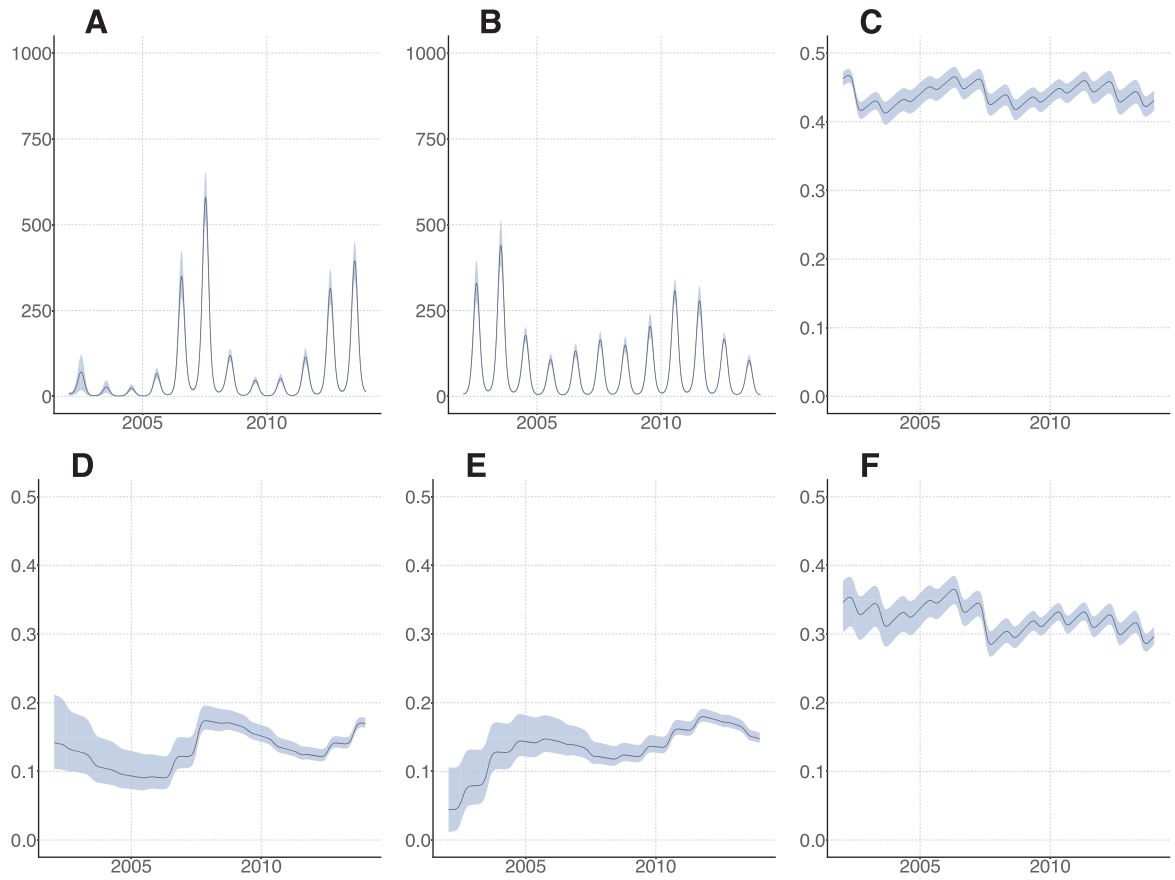


**Fig. 9. Number of infected and proportion of susceptible individuals.** Simulations without negative binomial noise using parameters from the MCMC chain associated with the highest posterior, calculated using both NDSS and DENFREE datasets. Posterior median (solid line) and 95% credible intervals (shaded blue area). A. Individuals infected with strain 1 in SEIR2psi model ($H_{I1} + H_{I21}$). B. Individuals infected with strain 2 in SEIR2psi model ($H_{I2} + H_{I12}$). C. Proportion of susceptible individuals in SEIR model ($H_S/N$). D. Proportion of susceptible individuals with immunity to strain 1 in SEIR2psi model ($H_{S1}/N$). E. Proportion of susceptible individuals with immunity to strain 2 in SEIR2psi model ($H_{S2}/N$). F. Proportion of individuals susceptible to both strains in SEIR2psi model ($H_S/N$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

models. These values are within the range observed in urban settings, suggesting that, despite very different population densities, the rural dynamics of dengue are not that dissimilar, or that dissimilarities are hidden by the variations between countries and populations and the uncertainties due to diverse estimation procedures, as this quantity is difficult to evaluate using incidence data. The median annual incidence at primary infection over the period is between 5 and 14, with large year-to-year variations. The estimated observation rate on surveillance data varies between models (9–13% in most models and 6–7% in the model with two interacting strains), indicating in both cases a high proportion of unreported infections, so that most of the transmission is due to unobserved infections. These values are in line with the large underrecognition highlighted in Wichmann et al. (2011), and also with other studies in South-East Asia (Graham et al., 1999; Thai et al., 2007; Yoon et al., 2012). These results are also coherent with a recent study emphasizing that more that 80% of dengue transmission is due to inapparent infections (ten Bosch et al., 2018).

This work has however several limitations. Firstly, model comparison was not straightforward between one strain and two strain models, both statistically and epidemiologically. From the statistical point of view, the differing number of observations between models led us to use simulation based-indicators. From the epidemiological point of view, single strain models, two strain models and observations on four serotypes may be hard to compare because some indicators cover different interpretations. For example, in single strain models, there is no distinction in the susceptible class between individuals immune to one strain only and naive ones, and there is no strain specific reproduction number or incidence.

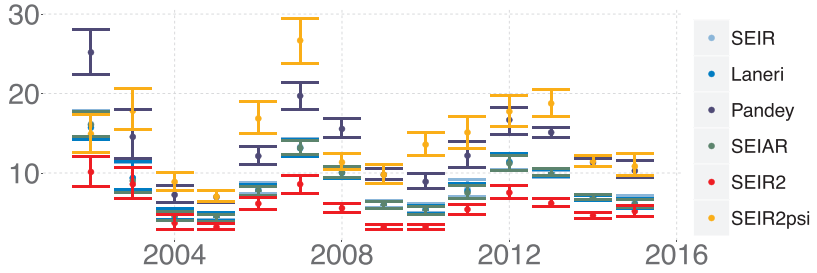


**Fig. 10. Annual incidence proportion of first dengue infection (%).** Median and credible intervals per year, over 2002–2015, based on simulations without negative binomial noise using parameters from the MCMC chain associated with the highest posterior, calculated using both NDSS and DENFREE datasets.

Secondly, the selected model formulations were restricted due to data availability. In particular, despite the endemicity of the four serotypes in rural Cambodia, we did not consider more than two dengue serotypes. This was done to limit model complexity, especially in the number of unknown initial conditions, but has also been previously shown to adequately describe dengue dynamics (Aguiar et al., 2013). When two serotypes were considered, we tested only interactions in terms of enhancement or restriction of infectiousness. We did not include models with (temporary) cross immunity, because of the too large increase in the number of parameters with respect to the data. We also did not include models with a finer spatial scale, even if small scale transmission plays a decisive role in dengue dynamics (Salje et al., 2012, 2017). On the one hand, NDSS data were only available at the district level, which was too large to follow transmission chains and too small for observing a sufficient number of cases. On the other hand, the clustered sampling protocol in the DENFREE study abnegated the interpretation of the spatial distribution of community cases. We also restricted the analysis to children under 15 years old and did not study the role of adults in transmission.

Thirdly, the projections are not completely able to describe the observed data, as most models overestimate the dengue epidemic in 2014. Nevertheless, 2014 was a particular year, with the lowest number of cases in the whole time series, potentially due to particular climatic conditions. In many countries in South-East Asia, except Malaysia, the reported incidence was lower than in 2013 (Cheng et al., 2017). Many provinces of Thailand also reported fewer cases than usual in 2014 (Reich et al., 2016). Our models are deterministic and do not take into account variations due to demographic stochasticity or environmental hazards such as climate.

Despite these limitations, combining two datasets permitted us to overcome some observation biases, such as the fact that surveillance data did not report serotype and DENFREE data did not reflect the long term dynamics. Nevertheless, some information is lacking in both datasets, in particular that on seroprevalence. Clearly the parameter estimations depend on the immunological status at the beginning of the simulations, especially in complex multi-strain models. As in our previous work (Champagne et al., 2016), our modeling study stresses the importance of seroprevalence data in order to more accurately estimate the initial conditions of our simulation and reduce identifiability problems. A seroprevalence survey in Cambodia would be of great value to evaluate the dengue burden, transmissibility potential and consider vaccination scenarios.

## 5. Conclusion

In conclusion, our analyses highlight that two-strain models with interacting effects better reproduce the long term dengue dynamics, but they are also difficult to parameterize relying on incidence data only. On the other hand, incorporating vector and asymptomatic components seems of limited added-value in this case when seasonality and underreporting are already accounted for. Although model complexity is framed by the scientific objectives, it must also be driven by the available data. The unavailability of mosquito data and the difficulty of observing asymptomatic infections questions their incorporation explicitly in the models. Another important aspect is related to the comparison of models considering the available data. In addition to model selection based on goodness of fit (Pandey et al., 2013), assessing the validity of model's outputs in terms of epidemiological features (such as reproduction numbers, annual incidence, dynamics of the susceptible classes) is an important step when data are scarce and identifiability issues are present.

## Acknowledgments

We thank the members of the Epidemiology and Public Health unit and the Virology unit in Institut Pasteur in Cambodia for their work during the DENFREE project. All the authors have been supported by a grant from Agence Nationale de la Recherche for the PANIC project (Targeting PAthogen's NIChe: a new approach for infectious diseases control in low-income countries: ANR-14-CE02-0015-01). The research leading to these results has also received funding from the European Commission Seventh Framework Program FP7 for the DENFREE project under Grant Agreement 282 378. CC and BC are partially supported by the “Pépiniere interdisciplinaire Eco-Evo-Devo” from the Centre National de la Recherche Scientifique (CNRS). CC is supported by the PhD studentship of the Groupe des Ecoles Nationales d’Economie et Statistique (GENES). The funders played no role in the study design, data collection, analysis, or preparation of the manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.epidem.2018.08.004.

## References

Aguiar, M., Ballesteros, S., Kooi, B.W., Stollenwerk, N., 2011. The role of seasonality and import in a minimalistic multi-strain dengue model capturing differences between primary and secondary infections: complex dynamics and its implications for data analysis. J. Theor. Biol. 289, 181–196. https://doi.org/10.1016/j.jtbi.2011.08.043.

Aguiar, M., Kooi, B.W., Rocha, F., Ghaffari, P., Stollenwerk, N., 2013. How much complexity is needed to describe the fluctuations observed in dengue hemorrhagic fever incidence data? Ecol. Complex. 16, 31–40. https://doi.org/10.1016/j.ecocom.2012.09.001. http://www.sciencedirect.com/science/article/pii/S1476945X12000670.

Aguiar, M., Paul, R., Sakuntabhai, A., Stollenwerk, N., 2014. Are we modelling the correct dataset? Minimizing false predictions for dengue fever in Thailand. Epidemiol. Infect. 142 (11), 2447–2459. https://doi.org/10.1017/S0950268813003348. http://journals.cambridge.org/article_S0950268813003348.

Aldstadt, J., Yoon, I.-K., Tannitisupawong, D., Jarman, R.G., Thomas, S.J., Gibbons, R.V., Uppapong, A., Iamsirithaworn, S., Rothman, A.L., Scott, T.W., Endy, T., 2012. Space-time analysis of hospitalised dengue patients in rural Thailand reveals important temporal intervals in the pattern of dengue virus transmission. Trop. Med. Int. Health: TM & IH 17 (9), 1076–1085. https://doi.org/10.1111/j.1365-3156.2012.03040.x.

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., Myers, M.F., George, D.B., Jaenisch, T., Wint, G.R.W., Simmons, C.P., Scott, T.W., Farrar, J.J., Hay, S.I., 2013. The global distribution and burden of dengue. Nature 496 (7446), 504–507. https://doi.org/10.1038/nature12060. https://www.nature.com/articles/nature12060.

Bretó, C., He, D., Ionides, E.L., King, A.A., 2009. Time series analysis via mechanistic models. Ann. Appl. Stat. 3 (1), 319–348. https://doi.org/10.1214/08-AOAS201. https://projecteuclid.org/euclid.aoas/1239888373.

Camacho, A., Ballesteros, S., Graham, A.L., Carrat, F., Ratmann, O., Cazelles, B., 2011. Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. Proc. R. Soc. Lond. B: Biol. Sci. https://doi.org/10.1098/rspb.2011.0300.. rspb20110300, http://rspb.royalsocietypublishing.org/content/early/2011/04/22/rspb.2011.0300.

R. Carnell, lhs: Latin hypercube samples. https://CRAN.R-project.org/package-lhs.

Cazelles, B., Cazelles, K., 2014. Major urban centers have weak influence on the timing of dengue epidemics in Southeast Asia. Int. J. Infect. Dis. 21 (Suppl. 1), 217. https://doi.org/10.1016/j.ijid.2014.03.873. http://www.sciencedirect.com/science/article/pii/S1201971214009321.

Champagne, C., Salthouse, D.G., Paul, R., Cao-Lormeau, V.-M., Roche, B., Cazelles, B., 2016. Structure in the variability of the basic reproductive number (R0) for Zika epidemics in the Pacific islands. eLife 5, e19874. https://doi.org/10.7554/eLife.19874. https://elifesciences.org/articles/19874.

Chan, M., Johansson, M.A., 2012. The incubation periods of Dengue viruses. PLOS ONE 7 (11), e50972. https://doi.org/10.1371/journal.pone.0050972.

Chareonsook, O., Foy, H.M., Teeraratkul, A., Silarug, N., 1999. Changing epidemiology of dengue hemorrhagic fever in Thailand. Epidemiol. Infect. 122 (1), 161–166.

Cheng, Q., Jing, Q., Spear, R.C., Marshall, J.M., Yang, Z., Gong, P., 2017. The interplay of climate, intervention and imported cases as determinants of the 2014 dengue outbreak in Guangzhou. PLOS Negl. Trop. Dis. 11 (6), e0005701. https://doi.org/10.1371/journal.pntd.0005701.

Clapham, H., Cummings, D.A.T., Nisalak, A., Kalayanarooj, S., Thaisomboonsuk, B., Klungthong, C., Fernandez, S., Srikiatkhachorn, A., Macareo, L.R., Lessler, J., Reiser, J., Yoon, I.-K., 2015. Epidemiology of infant dengue cases illuminates serotype-specificity in the interaction between immunity and disease, and changes in transmission dynamics. PLOS Negl. Trop. Dis. 9 (12), e0004262. https://doi.org/10.1371/journal.pntd.0004262.

Coudeville, L., Garnett, G.P., 2012. Transmission dynamics of the four dengue serotypes in Southern Vietnam and the potential impact of vaccination. PLOS ONE 7 (12), e51244. https://doi.org/10.1371/journal.pone.0051244.

Cuong, H.Q., Vu, N.T., Cazelles, B., Boni, M.F., Thai, K.T., Rabaa, M.A., Quang, L.C., Simmons, C.P., Huu, T.N., Anders, K.L., 2013. Spatiotemporal dynamics of dengue epidemics, Southern Vietnam. Emerg. Infect. Dis. 19 (6), 945–953. https://doi.org/

10.3201/eid1906.121323. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3713821/.
Diekmann, O., Heesterbeek, J.A.P., Roberts, M.G., 2010. The construction of next-generation matrices for compartmental epidemic models. J. R. Soc. Interface 7 (47), 873–885. https://doi.org/10.1098/rsif.2009.0386. http://rsif.royalsocietypublishing.org/content/7/47/873.
Duong, V., Lambrechts, L., Paul, R.E., Ly, S., Lay, R.S., Long, K.C., Huy, R., Tarantola, A., Scott, T.W., Sakuntabhai, A., Buchy, P., 2015. Asymptomatic humans transmit dengue virus to mosquitoes. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1508114112.
Dureau, J., Ballesteros, S., Bogich, T., 2013. ssm: Inference for State Space Models like playing with duplo blocks, original-date: 2014-08-29T13:15:10Z (Nov.). https://github.com/JDureau/ssm.
Dye, C., Williams, B., 1995. Nonlinearities in the dynamics of indirectly-transmitted infections (or, does having a vector make a difference? In: Grenfell, B.T., Grenfell, B.T., Dobson, A.P., Dobson, A.P. (Eds.), Ecology of Infectious Diseases in Natural Populations. Cambridge University Press Edition, pp. 260–279 http://www.core/books/ecology-of-infectious-diseases-in-natural-populations/nonlinearities-in-the-dynamics-of-indirectly-transmitted-infections-or-does-having-a-vector-make-a-difference/1E571DC1187028C2262A1D6EFE5A73A4.
Evans, N., Chapman, M., Chappell, M., Godfrey, K., 2002. The structural identifiability of a general epidemic (SIR) model with seasonal forcing. IFAC Proc. Vol. 35 (1), 109–114. https://doi.org/10.3182/20020721-6-ES-1901.01327. http://linkinghub.elsevier.com/retrieve/pii/S1474667015397482.
Ferguson, N., Anderson, R., Gupta, S., 1999. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. Proc. Natl. Acad. Sci. USA 96 (2), 790–794. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC15215/.
Graham, R.R., Juffrie, M., Tan, R., Hayes, C.G., Laksono, I., Ma'roef, C., Erlin, n., Sutaryo, n., Porter, K.R., Halstead, S.B., 1999. A prospective seroepidemiologic study on dengue in children four to nine years of age in Yogyakarta, Indonesia I. studies in 1995–1996. Am. J. Trop. Med. Hyg. 61 (3), 412–419.
Grange, L., Simon-Loriere, E., Sakuntabhai, A., Gresh, L., Paul, R., Harris, E., 2014. Epidemiological risk factors associated with high:global frequency of inapparent dengue virus: infections. Front. Immunol. 5https://doi.org/10.3389/fimmu.2014.00280. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4052743/.
Guha-Sapir, D., Schimmer, B., 2005. Dengue fever: new paradigms for a changing epidemiology. Emerg. Themes Epidemiol. 2, 1. https://doi.org/10.1186/1742-7622-2-1.
Guzman, M.G., Halstead, S.B., Artsob, H., Buchy, P., Farrar, J., Gubler, D.J., Hunsperger, E., Kroeger, A., Margolis, H.S., Martínez, E., Nathan, M.B., Pelegrino, J.L., Simmons, C., Yoksan, S., Peeling, R.W., 2010. Dengue: a continuing global threat. Nat. Rev. Microbiol. 8, S7–S16. https://doi.org/10.1038/nrmicro2460. http://www.nature.com/nrmicro/journal/v8/n12_supp/full/nrmicro2460.html.
Halstead, S.B., 2007. Dengue. Lancet (London, England) 370 (9599), 1644–1652. https://doi.org/10.1016/S0140-6736(07)61687-0.
Heesterbeek, H., Anderson, R.M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., Eames, K.T.D., Edmunds, W.J., Frost, S.D.W., Funk, S., Hollingsworth, T.D., House, T., Isham, V., Klepac, P., Lessler, J., Lloyd-Smith, J.O., Metcalf, C.J.E., Mollison, D., Pellis, L., Pulliam, J.R.C., Roberts, M.G., Viboud, C., Isaac Newton Institute IDD Collaboration, 2015. Modeling infectious disease dynamics in the complex landscape of global health. Science 347 (6227), aaa4339. https://doi.org/10.1126/science.aaa4339.
Huy, R., Buchy, P., Conan, A., Ngan, C., Ong, S., Ali, R., Duong, V., Yit, S., Ung, S., Te, V., Chroeung, N., Pheaktra, N.C., Uok, V., Vong, S., 2010. National dengue surveillance in Cambodia 1980–2008: epidemiological and virological trends and the impact of vector control. Bull. World Health Organ. 88 (9), 650–657. https://doi.org/10.2471/BLT.09.073908. http://www.who.int/bulletin/volumes/88/9/09-073908.pdf.
Imai, N., Dorigatti, I., Cauchemez, S., Ferguson, N.M., 2015. Estimating dengue transmission intensity from sero-prevalence surveys in multiple countries. PLoS Negl. Trop. Dis. 9 (4). https://doi.org/10.1371/journal.pntd.0003719. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400108/.
Imai, N., Dorigatti, I., Cauchemez, S., Ferguson, N.M., 2016. Estimating dengue transmission intensity from case-notification data from multiple countries. PLOS Negl. Trop. Dis. 10 (7), e0004833. https://doi.org/10.1371/journal.pntd.0004833.
Johansson, M.A., Hombach, J., Cummings, D.A., 2011. Models of the impact of dengue vaccines: a review of current research and potential approaches. Vaccine 29 (35), 5860–5868. https://doi.org/10.1016/j.vaccine.2011.06.042. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4327892/.
Kao, Y.-H., Eisenberg, M.C., 2018. Practical unidentifiability of a simple vector-borne disease model: implications for parameter estimation and intervention assessment. Epidemics. https://doi.org/10.1016/j.epidem.2018.05.010. http://www.sciencedirect.com/science/article/pii/S1755436517301627.
King, A.A., Ionides, E.L., Pascual, M., Bouma, M.J., 2008. Inapparent infections and cholera dynamics. Nature 454 (7206), 877–880. https://doi.org/10.1038/nature07084. http://www.nature.com/doifinder/10.1038/nature07084.
Laneri, K., Bhadra, A., Ionides, E.L., Bouma, M., Dhiman, R.C., Yadav, R.S., Pascual, M., 2010. Forcing versus feedback: epidemic malaria and monsoon rains in Northwest India. PLoS Comput. Biol. 6 (9), e1000898. https://doi.org/10.1371/journal.pcbi.1000898.
L'Azou, M., Moureau, A., Sarti, E., Nealon, J., Zambrano, B., Wartel, T.A., Villar, L., Capeding, M.R., Ochiai, R.L., 2016. Symptomatic dengue in children in 10 Asian and Latin American Countries. New Engl. J. Med. 374 (12), 1155–1166. https://doi.org/10.1056/NEJMoa1503877.
L'Azou, M., Assoukpa, J., Fanouillere, K., Plennevaux, E., Bonaparte, M., Bouckenooghe, A., Frago, C., Noriega, F., Zambrano, B., Ochiai, R.L., Guy, B., Jackson, N., 2018. Dengue seroprevalence: data from the clinical development of a tetravalent dengue vaccine in 14 countries (2005–2014). Trans. R. Soc. Trop. Med. Hyg. https://doi.org/10.1093/trstmh/try037.
Liu-Helmersson, J., Stenlund, H., Wilder-Smith, A., Rocklöv, J., 2014. Vectorial Capacity of *Aedes aegypti*: effects of temperature and implications for global dengue epidemic potential. PLOS ONE 9 (3), e89783. https://doi.org/10.1371/journal.pone.0089783.
Mammen Jr., M.P., Pimgate, C., Koenraadt, C.J., Rothman, A.L., Aldstadt, J., Nisalak, A., Jarman, R.G., Jones, J.W., Srikiatkhachorn, A., Ypil-Butac, C.A., et al., 2008. Spatial and temporal clustering of dengue virus transmission in Thai villages. PLoS Med. 5 (11), e205. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0050205.
Messina, J.P., Brady, O.J., Scott, T.W., Zou, C., Pigott, D.M., Duda, K.A., Bhatt, S., Katzelnick, L., Howes, R.E., Battle, K.E., Simmons, C.P., Hay, S.I., 2014. Global spread of dengue virus types: mapping the 70 year history. Trends Microbiol. 22 (3), 138–146. https://doi.org/10.1016/j.tim.2013.12.011.
Messina, J.P., Brady, O.J., Pigott, D.M., Golding, N., Kraemer, M.U.G., Scott, T.W., Wint, G.R.W., Smith, D.L., Hay, S.I., 2015. The many projected futures of dengue. Nat. Rev. Microbiol. 13 (4), 230–239. https://doi.org/10.1038/nrmicro3430.
Miao, H., Xia, X., Perelson, A.S., Wu, H., 2011. On identifiability of nonlinear ODE models and applications in viral dynamics. SIAM Rev. Soc. Ind. Appl. Math. 53 (1), 3–39. https://doi.org/10.1137/090757009. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3140286/.
Muhammad Azami, N.A., Salleh, S.A., Neoh, H.-m., Syed Zakaria, S.Z., Jamal, R., 2011. Dengue epidemic in Malaysia: not a predominantly urban disease anymore. BMC Res. Notes 4, 216. https://doi.org/10.1186/1756-0500-4-216.
Nguyen, N.M., Kien, D.T.H., Tuan, T.V., Quyen, N.T.H., Tran, C.N.B., Thi, L.V., Thi, D.L., Nguyen, H.L., Farrar, J.J., Holmes, E.C., Rabaa, M.A., Bryant, J.E., Nguyen, T.T., Nguyen, H.T.C., Nguyen, L.T.H., Pham, M.P., Nguyen, H.T., Luong, T.T.H., Wills, B., Nguyen, C.V.V., Wolbers, M., Simmons, C.P., 2013. Host and viral features of human dengue cases shape the population of infected and infectious *Aedes aegypti* mosquitoes. Proc. Natl. Acad. Sci. 110 (22), 9072–9077. https://doi.org/10.1073/pnas.1303395110. http://www.pnas.org/content/110/22/9072.
Nisalak, A., Clapham, H.E., Kalayanarooj, S., Klungthong, C., Thaisomboonsuk, B., Fernandez, S., Reiser, J., Srikiatkhachorn, A., Macareo, L.R., Lessler, J.T., Cummings, D.A.T., Yoon, I.-K., 2016. Forty years of dengue surveillance at a Tertiary Pediatric Hospital in Bangkok, Thailand, 1973–2012. Am. J. Trop. Med. Hyg. 94 (6), 1342–1347. https://doi.org/10.4269/ajtmh.15-0337.
Ong, J., Yap, G., Hapuarachchi, C., Tang, C.S., Liang, S., Rajarethinam, J., Tan, S., Koo, C., Zeng, K., Ycasas, J., Tay, L., Chong, C.-S., Abdul Razak, M.A., Lee, C., Chiang, S., Lai, Y.-L., Ng, L.-C., 2017. Dengue review of 2016 and outlook for 2017. Singap. Epidemiol. News Bull. 43 (2), 47–56.
National Institute of Statistics, Ministry of Planning, 2009. Phnom Penh, Cambodia, General Population Census of Cambodia 2008. National Report on Final Census Results.
Pandey, A., Mubayi, A., Medlock, J., 2013. Comparing vector-host and SIR models for dengue transmission. Math. Biosci. 246 (2), 252–259. https://doi.org/10.1016/j.mbs.2013.10.007. http://linkinghub.elsevier.com/retrieve/pii/S0025556413002435.
Perkins, T.A., Paz-Soldan, V.A., Stoddard, S.T., Morrison, A.C., Forshey, B.M., Long, K.C., Halsey, E.S., Kochel, T.J., Elder, J.P., Kitron, U., Scott, T.W., Vazquez-Prokopec, G.M., 2016. Calling in sick: impacts of fever on intra-urban human mobility. Proc. R. Soc. B 283 (1834), 20160390. https://doi.org/10.1098/rspb.2016.0390. http://rspb.royalsocietypublishing.org/content/283/1834/20160390.
Prayitno, A., Taurel, A.-F., Nealon, J., Satari, H.I., Karyanti, M.R., Sekartini, R., Soedjatmiko, S., Gunardi, H., Medise, B.E., Sasmono, R.T., Simmerman, J.M., Bouckenooghe, A., Hadinegoro, S.R., 2017. Dengue seroprevalence and force of primary infection in a representative population of urban dwelling Indonesian children. PLOS Negl. Trop. Dis. 11 (6), e0005621. https://doi.org/10.1371/journal.pntd.0005621.
Reich, N.G., Shrestha, S., King, A.A., Rohani, P., Lessler, J., Kalayanarooj, S., Yoon, I.-K., Gibbons, R.V., Burke, D.S., Cummings, D.A.T., 2013. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. J. R. Soc. Interface 10 (86), 20130414. https://doi.org/10.1098/rsif.2013.0414.
Reich, N.G., Lauer, S.A., Sakrejda, K., Iamsirithaworn, S., Hinjoy, S., Suangtho, P., Suthachana, S., Clapham, H.E., Salje, H., Cummings, D.A.T., Lessler, J., 2016. Challenges in real-time prediction of infectious disease: a case study of dengue in Thailand. PLOS Negl. Trop. Dis. 10 (6), e0004761. https://doi.org/10.1371/journal.pntd.0004761.
Reiner, R.C., Perkins, T.A., Barker, C.M., Niu, T., Chaves, L.F., Ellis, A.M., George, D.B., Le Menach, A., Pulliam, J.R.C., Bisanzio, D., Buckee, C., Chiyaka, C., Cummings, D.A.T., Garcia, A.J., Gatton, M.L., Gething, P.W., Hartley, D.M., Johnston, G., Klein, E.Y., Michael, E., Lindsay, S.W., Lloyd, A.L., Pigott, D.M., Reisen, W.K., Ruktanonchai, N., Singh, B.K., Tatem, A.J., Kitron, U., Hay, S.I., Scott, T.W., Smith, D.L., 2013. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. J. R. Soc. Interface 10 (81), 20120921. https://doi.org/10.1098/rsif.2012.0921.
Reller, M.E., Bodinayake, C., Nagahawatte, A., Devasiri, V., Kodikara-Arachichi, W., Strouse, J.J., Broadwater, A., Østbye, T., de Silva, A., Woods, C.W., 2012. Unsuspected dengue and acute febrile illness in rural and semi-urban southern Sri Lanka. Emerg. Infect. Dis. 18 (2), 256–263. https://doi.org/10.3201/eid1802.110962. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3310451/.
Rocha, F., Aguiar, M., Souza, M., Stollenwerk, N., 2013. Time-scale separation and centre manifold analysis describing vector-borne disease dynamics. Int. J. Comput. Math. 90 (10), 2105–2125. https://doi.org/10.1080/00207160.2013.783208.
Rocha, F., Mateus, L., Skwara, U., Aguiar, M., Stollenwerk, N., 2016. Understanding dengue fever dynamics: a study of seasonality in vector-borne disease models. Int. J. Comput. Math. 93 (8), 1405–1422. https://doi.org/10.1080/00207160.2015.

1050961.
Salje, H., Lessler, J., Endy, T.P., Curriero, F.C., Gibbons, R.V., Nisalak, A., Nimmannitya, S., Kalayanarooj, S., Jarman, R.G., Thomas, S.J., Burke, D.S., Cummings, D.A.T., 2012. Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. Proc. Natl. Acad. Sci. USA 109 (24), 9535–9538. https://doi.org/10.1073/pnas.1120621109.
Salje, H., Lessler, J., Berry, I.M., Melendrez, M.C., Endy, T., Kalayanarooj, S., A-Nuegoonpipat, A., Chanama, S., Sangkijporn, S., Klungthong, C., Thaisomboonsuk, B., Nisalak, A., Gibbons, R.V., Iamsirithaworn, S., Macareo, L.R., Yoon, I.-K., Sangarsang, A., Jarman, R.G., Cummings, D.A.T., 2017. Dengue diversity across spatial and temporal scales: local structure and the effect of host population size. Science 355 (6331), 1302–1306. https://doi.org/10.1126/science.aaj9384. http://science.sciencemag.org/content/355/6331/1302.
Sharp, T.M., Tomashek, K.M., Read, J.S., Margolis, H.S., Waterman, S.H., 2017. A new look at an old disease: recent: insights into the global epidemiology of dengue. Curr. Epidemiol. Rep. 4 (1), 11–21. https://doi.org/10.1007/s40471-017-0095-y.
Shepard, D.S., Undurraga, E.A., Halasa, Y.A., Stanaway, J.D., 2016. The global economic burden of dengue: a systematic analysis. Lancet Infect. Dis. 16 (8), 935–941. https://doi.org/10.1016/S1473-3099(16)00146-8.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 64 (4), 583–639. https://doi.org/10.1111/1467-9868.00353.
Stoddard, S.T., Forshey, B.M., Morrison, A.C., Paz-Soldan, V.A., Vazquez-Prokopec, G.M., Astete, H., Reiner, R.C., Vilcarromero, S., Elder, J.P., Halsey, E.S., Kochel, T.J., Kitron, U., Scott, T.W., 2013. House-to-house human movement drives dengue virus transmission. Proc. Natl. Acad. Sci. USA 110 (3), 994–999. https://doi.org/10.1073/pnas.1213349110. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3549073/.
Strickman, D., Sithiprasasna, R., Kittayapong, P., Innis, B.L., 2000. Distribution of dengue and Japanese encephalitis among children in rural and suburban Thai villages. Am. J. Trop. Med. Hyg. 63 (1–2), 27–35.
Team, R.D.C., 2015. R: A Language and Environment for Statistical Computing. http://www.R-project.org.
ten Bosch, Q.A., Singh, B.K., Hassan, M.R.A., Chadee, D.D., Michael, E., 2016. The role of serotype interactions and seasonality in dengue model selection and control: insights from a pattern matching approach. PLOS Negl. Trop. Dis. 10 (5), e0004680. https://doi.org/10.1371/journal.pntd.0004680.
ten Bosch, Q.A., Clapham, H.E., Lambrechts, L., Duong, V., Buchy, P., Althouse, B.M., Lloyd, A.L., Waller, L.A., Morrison, A.C., Kitron, U., Vazquez-Prokopec, G.M., Scott, T.W., Perkins, T.A., 2018. Contributions from the silent majority dominate dengue virus transmission. PLoS Pathogens 14 (5). https://doi.org/10.1371/journal.ppat.1006965. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5933708/.
Teurlai, M., Huy, R., Cazelles, B., Duboz, R., Baehr, C., Vong, S., 2012. Can human movements explain heterogeneous propagation of dengue fever in Cambodia? PLOS Negl. Trop. Dis. 6 (12), e1957. https://doi.org/10.1371/journal.pntd.0001957.
Thai, K.T., Binh, T.Q., Giao, P.T., Phuong, H.L., Hung, L.Q., Nam, N.V., Nga, T.T., Groen, J., Nagelkerke, N., Vries, P.J., 2005. Seroprevalence of dengue antibodies, annual incidence and risk factors among children in southern Vietnam. Trop. Med. Int. Health 10 (4), 379–386. https://doi.org/10.1111/j.1365-3156.2005.01388.x/full.
Thai, K.T.D., Nga, T.T.T., Van Nam, N., Phuong, H.L., Giao, P.T., Hung, L.Q., Binh, T.Q., Van Doornum, G.J.J., De Vries, P.J., 2007. Incidence of primary dengue virus infections in Southern Vietnamese children and reactivity against other flaviviruses. Trop. Med. Int. Health 12 (12), 1553–1557. https://doi.org/10.1111/j.1365-3156.2007.01964.x.
Tien, N.T.K., Luxemburger, C., Toan, N.T., Pollissard-Gadroy, L., Huong, V.T.Q., Van Be, P., Rang, N.N., Wartel, T.-A., Lang, J., 2010. A prospective cohort study of dengue infection in schoolchildren in Long Xuyen, Viet Nam. Trans. R. Soc. Trop. Med. Hyg. 104 (9), 592–600. https://doi.org/10.1016/j.trstmh.2010.06.003.
Vong, S., Khieu, V., Glass, O., Ly, S., Duong, V., Huy, R., Ngan, C., Wichmann, O., Letson, G.W., Margolis, H.S., Buchy, P., 2010. Dengue incidence in urban and rural Cambodia: results from population-based active fever surveillance, 2006–2008. PLoS Negl. Trop. Dis. 4 (11), e903. https://doi.org/10.1371/journal.pntd.0000903.
Vongpunsawad, S., Intharasongkroh, D., Thongmee, T., Poovorawan, Y., 2017. Seroprevalence of antibodies to dengue and chikungunya viruses in Thailand. PLOS ONE 12 (6), e0180560. https://doi.org/10.1371/journal.pone.0180560.
WHO, 2017. Dengue – World Health Organisation. http://www.who.int/denguecontrol/en/.
Wichmann, O., Yoon, I.-K., Vong, S., Limkittikul, K., Gibbons, R.V., Mammen, M.P., Ly, S., Buchy, P., Sirivichayakul, C., Buathong, R., Huy, R., Letson, G.W., Sabchareon, A., 2011. Dengue in Thailand and Cambodia: an assessment of the degree of under-recognized disease burden based on reported cases. PLoS Negl. Trop. Dis. 5 (3), e996. https://doi.org/10.1371/journal.pntd.0000996.
Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. http://ggplot2.org.
Yoon, I.-K., Rothman, A.L., Tannitisupawong, D., Srikiatkhachorn, A., Jarman, R.G., Aldstadt, J., Nisalak, A., Mammen, M.P., Thammapalo, S., Green, S., Libraty, D.H., Gibbons, R.V., Getis, A., Endy, T., Jones, J.W., Koenraadt, C.J.M., Morrison, A.C., Fansiri, T., Pimgate, C., Scott, T.W., 2012. Underrecognized mildly symptomatic viremic dengue virus infections in rural Thai schools and villages. J. Infect. Dis. 206 (3), 389–398. https://doi.org/10.1093/infdis/jis357. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3490697/.