Dengue modeling in rural Cambodia: statistical performance versus epidemiological relevance

Clara CHAMPAGNE *[1,2], Richard PAUL[3,4], Sowath LY[6], Veasna DUONG[5], Rithea LEANG[7] and Bernard CAZELLES †[1,3,8]

[1]Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Research University, 75005 Paris, France

[2]CREST, ENSAE, Université Paris Saclay, 5 avenue Henry Le Chatelier, 91120 Palaiseau, France

[3]Institut Pasteur, Unité de Génétique Fonctionnelle des Maladies Infectieuses, Department of Genomes and Genetics, F-75724 Paris cedex 15, France

[4]Centre National de la Recherche Scientifique, Unité de Recherche associée 3012, Paris, France

[5]Institut Pasteur in Cambodia, Virology Unit, Phnom Penh, Cambodia

[6]Institut Pasteur in Cambodia, Epidemiology and Public Health Unit, Phnom Penh, Cambodia

[7]National Dengue Control Program, Cambodia

[8]International Center for Mathematical and Computational Modeling of Complex Systems (UMMISCO), UMI 209 UPMC/IRD, Bondy cedex, France

October 25, 2017

*champagn@biologie.ens.fr

†cazelles@biologie.ens.fr

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, but also by studying the coherence of epidemiological characteristics (reproduction numbers, incidence proportion, dynamics of the susceptible class) in each model. Considering the available data, our analyses on transmission dynamics in a rural endemic setting highlight both the importance of using two-strain models with interacting effects and the lack of added value of incorporating vector and explicit asymptomatic components.

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 [1]. Dengue is ubiquitous in the tropics and the subtropics, particularly in Southeast Asia, the Pacific and the Americas [2].

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 [3, 4], an estimated 390 million infections every year and about 50-100 million symptomatic cases worldwide and a high disease burden [5, 6]. Nowadays, there are more cases of dengue worldwide than any other arboviral disease [7, 8, 9].

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 [10, 11]. 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 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 [12, 13, 14, 15]. Few studies have been carried out in rural settings [16, 17, 18], despite growing evidence that rural dengue is an increasing problem. Guha-Sapir and Schimmer [19] observed shifts in modal age, rural spread, and social determinants of dengue susceptibility, with major implications for health services. Muhammad Azami et al. [20] observed similar dengue seroprevalence rates between urban and rural samples, showing that dengue is not confined to urban areas in Malaysia. Chareonsook et al. [21] 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 [22, 23]. 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 [24, 25, 26].

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 [27, 28, 13, 29]. 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 120km northeast from the capital Phnom Penh. Dengue is endemic and strongly seasonal (cf. Figure 1), with outbreaks occuring 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 30km 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).

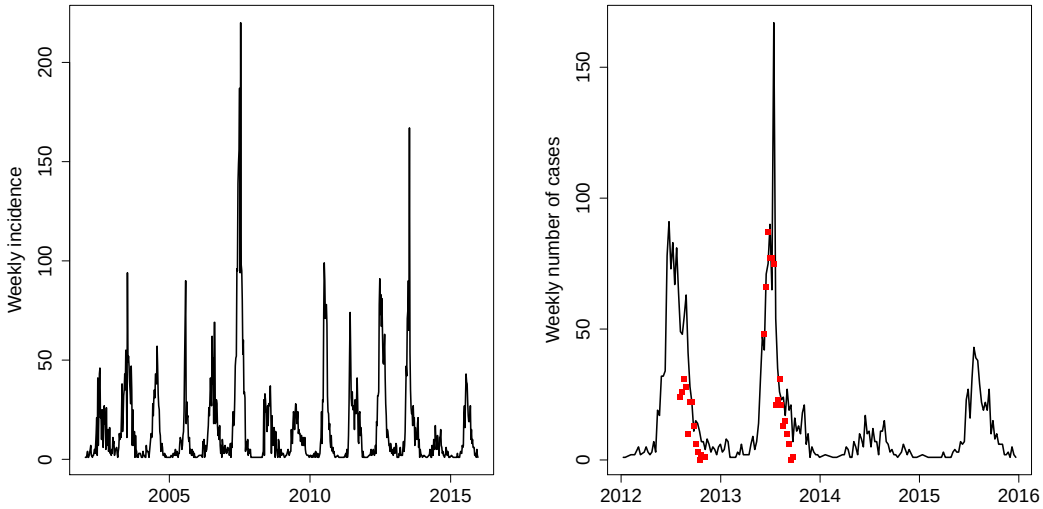

Figure 1 – **a) NDSS data.** Weekly number of cases in the four districts of Kampong Cham province. **b) NDSS and DENFREE data.** The black line gives the series of NDSS cases per week in the four districts of the Kampong Cham province. The red dots are the DENFREE cases per week in the DENFREE study area.

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 and DENV infection was confirmed by qRT-PCR. Positive DENV cases were considered as index cases, and an outbreak investigation was initiated in their neighbourhood, in order to detect asymptomatic or mildly symptomatic cases. The study protocol is extensively detailed in Duong et al. [30].

We used the series of the total number of cases per week (index cases and outbreak investigation cases) between August 6th (first week when more than 100 people were tested

for dengue in outbreak investigation) and November 5th 2012 and between June 10th (first week with outbreak investigation) and September 23rd 2013. 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, and it allowed a comparison with other dengue reporting systems in Cambodia, which are mainly done at paedriatric hospitals. Information on the serotype responsible for infection, and symptomatic/asymptomatic status of the patients were available.

Children under 15 years old	**2012**	**2013**
Observed symptomatic cases	189	571
Denv-1	183	448
Denv-2, Denv-3, Denv-4	1	122
Observed asymptomatic cases	4	28
Children tested in the community	1722	4119

Table 1 – **Number of cases (index cases and community cases) under 15 years old, in the DENFREE study.** Cases are restricted to the period 6th Aug - 5th Nov 2012 and 10th Jun - 23rd Sept 2013. During the whole study, 236 symptomatic cases were collected in 2012 and 574 in 2013. Serotype was unknown for 5 cases in 2012. One case in 2013 was coinfected with DENV-1 and DENV-2 and was not included in models with two strains (serotypes).

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)[31, 24], involving the paediatric departments of several hospitals throughout the country. 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 mis-diagnosis 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 (cf. Figure 1). 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. As displayed in figure 1, NDSS and DENFREE data have similar dynamics.

DENFREE	**NDSS**
Observations in 2012-2013	Observations in 2002-2015
No observation during inter-epidemic period, dataset starts during the epidemic peak	Observations all year round
Laboratory confirmation	Clinical diagnosis
Observation of non hospitalized cases including some asymptomatic infections	Hospital cases only
Clustered collecting process	Stable reporting process over time
An observation rate can be estimated	Unknown observation rate
Known serotype	Unknown serotype

Table 2 – **Comparison of both datasets**

2.1.4 Population

We take as the reference population (N=161391) 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 [32]. Since the DENFREE study was conducted in a subpart of this area, we calculated the total population for the DENFREE study (n=65208) as the sum of the population of children under 15 years old in all the villages investigated in either 2012 or 2013 [32].

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. Figure 2). In this model, the basic reproduction number, i.e. the number of secondary human infections resulting from the introduction of a single infected individual in an entirely susceptible population, is $R_0^{SEIR}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$.

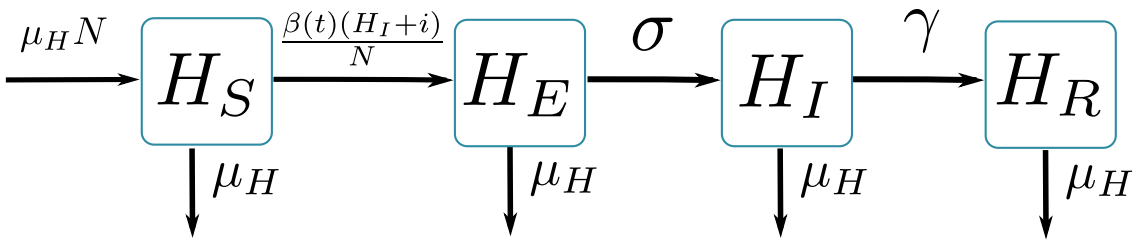

Figure 2 – **Graphical representation of SEIR model.** H_S susceptible individuals; H_E infected (not yet infectious) individuals; H_I infectious individuals; H_R recovered individuals; $\beta(t)$ is the transmission parameter; σ is the rate at which H_E-individuals move to the infectious class H_I; infectious individuals (H_I) then recover at rate γ; individuals leave the children population at rate μ_H. $H_S + H_E + H_I + H_R = N$.

$$\begin{aligned}
\frac{dH_S}{dt} &= \mu_H N - \beta(t)\frac{(H_I+i)H_S}{N} - \mu_H H_S \\
\frac{dH_E}{dt} &= \beta(t)\frac{(H_I+i)H_S}{N} - \sigma H_E - \mu_H H_E \\
\frac{dH_I}{dt} &= \sigma H_E - \gamma H_I - \mu_H H_I \\
\frac{dH_R}{dt} &= \gamma H_I - \mu_H H_R
\end{aligned} \tag{1}$$

This model is compared with two other models that include the mosquito vector transmission components. In the first one, derived from Pandey et al. [29], the vector is modelled explicitly with three compartments (Susceptible-Exposed-Infected) (cf. Figure 3). In the second one, derived from Laneri et al.[33], the vector is modelled implicitly as an

external force of infection including two stages, latent (κ) and current (λ) (cf. Figure 4). We derived R_0 for each model as $R_0^{Pandey}(t) = \frac{\beta_H \beta_V(t) \sigma \tau}{(\gamma+\mu_H)(\sigma+\mu_H)\mu_V(\mu_V+\tau)}$ and the estimation $R_0^{Laneri}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$ [34]. 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 individual in a entirely susceptible population), and not the reproduction ratio per generation provided through the use of the next generation matrix.

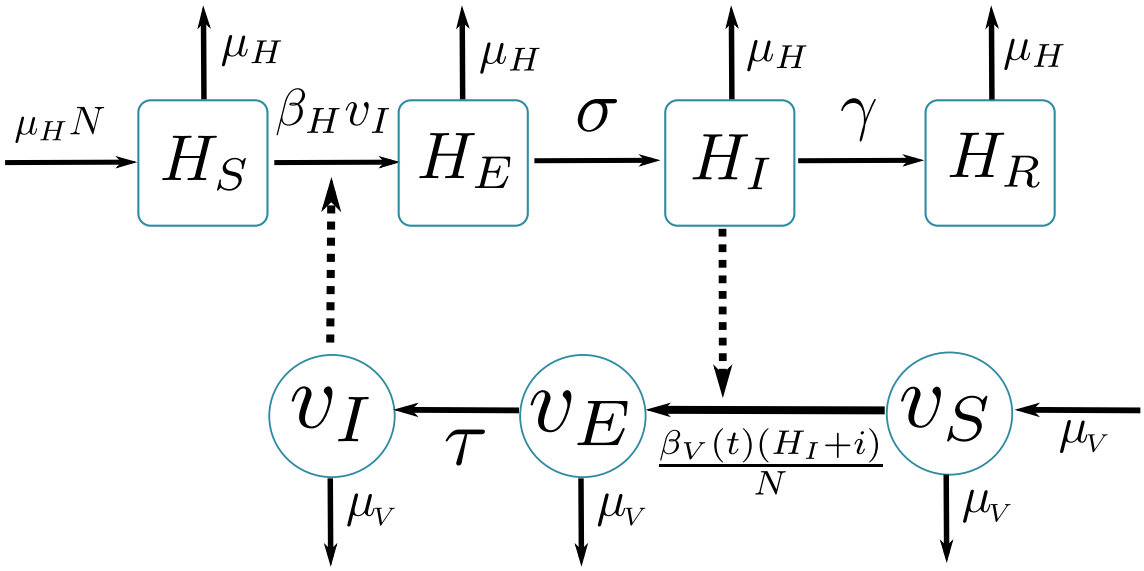

Figure 3 – **Graphical representation of Pandey model [29].** Squared boxes and circles correspond respectively to human and vector compartments. Plain arrows represent transitions from one state to the next. Dashed arrows indicate interactions between humans and vectors. H_S susceptible individuals; H_E infected (not yet infectious) individuals; H_I infectious individuals; H_R recovered individuals; β_H is the transmission parameter from vector to human; σ is the rate at which H_E-individuals move to the infectious class H_I; infectious individuals (H_I) then recover at rate γ; individuals leave the children population at rate μ_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; τ is the rate at which v_E-vectors move to the infectious class v_I; vectors die at rate μ_V.

The equations describing the Pandey model are:

$$
\begin{aligned}
\frac{dH_S}{dt} &= \mu_H N - \beta_H v_I H_S - \mu_H H_S \\
\frac{dH_E}{dt} &= \beta_H v_I H_S - \sigma H_E - \mu_H H_E \\
\frac{dH_I}{dt} &= \sigma H_E - \gamma H_I - \mu_H H_I \\
\frac{dH_R}{dt} &= \gamma H_I - \mu_H H_R \\
\frac{dv_S}{dt} &= \mu_V - \beta_V(t)\frac{(H_I + i)}{N} v_S - \mu_V v_S \\
\frac{dv_E}{dt} &= \beta_V(t)\frac{(H_I + i)}{N} v_S - \tau v_E - \mu_V v_E \\
\frac{dv_I}{dt} &= \tau v_E - \mu_V v_I
\end{aligned} \tag{2}
$$

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

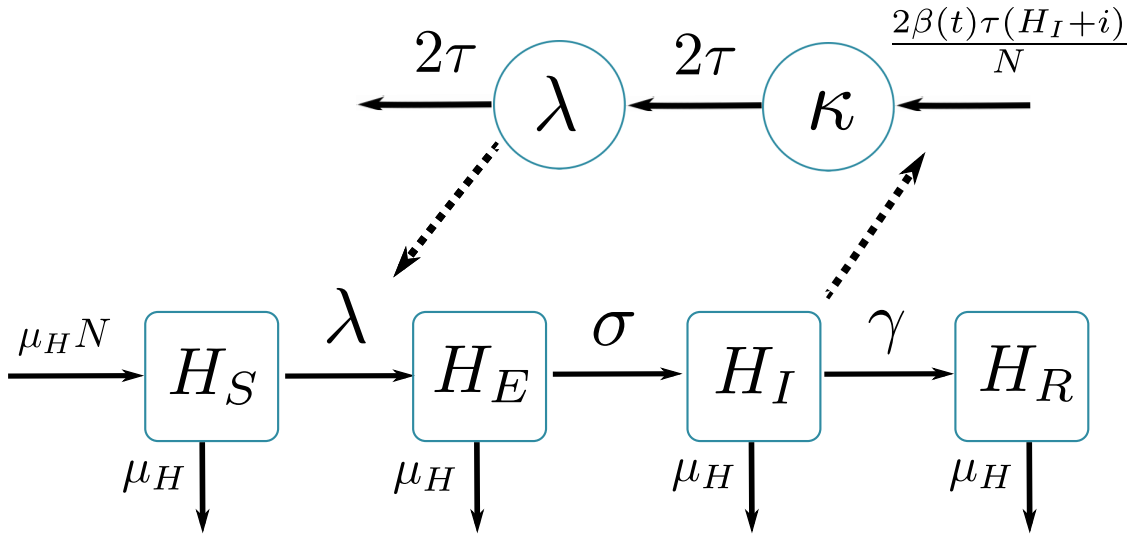

Figure 4 – **Graphical representation of Laneri model [33].** Squared boxes and circles correspond respectively to human and vector compartments. Plain arrows represent transitions from one state to the next. Dashed arrows indicate interactions between humans and vectors. H_S susceptible individuals; H_E infected (not yet infectious) individuals; H_I infectious individuals; H_R recovered individuals; σ is the rate at which H_E-individuals move to the infectious class H_I; infectious individuals (H_I) then recover at rate γ; individuals leave the children population at rate μ_H; $H_S + H_E + H_I + H_R = N$; implicit vector-borne transmission is modelled with the compartments κ and λ; λ current force of infection; κ latent force of infection reflecting the exposed state for mosquitoes during the extrinsic incubation period; $\beta(t)$ is the transmission parameter; τ is the transition rate associated with the extrinsic incubation period.

159 The equations describing the Laneri model are:

$$
\begin{aligned}
\frac{dH_S}{dt} &= \mu_H N - \lambda H_S - \mu_H H_S \\
\frac{dH_E}{dt} &= \lambda H_S - \sigma H_E - \mu_H H_E \\
\frac{dH_I}{dt} &= \sigma H_E - \gamma H_I - \mu_H H_I \\
\frac{dH_R}{dt} &= \gamma H_I - \mu_H H_R \\
\frac{d\kappa}{dt} &= \beta(t)\frac{2(H_I + i)\tau}{N} - 2\tau\kappa \\
\frac{d\lambda}{dt} &= 2\tau\kappa - 2\tau\lambda
\end{aligned}
\tag{3}
$$

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. Figure 5). 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 [30], and therefore, $R_0^{SEIAR}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$.

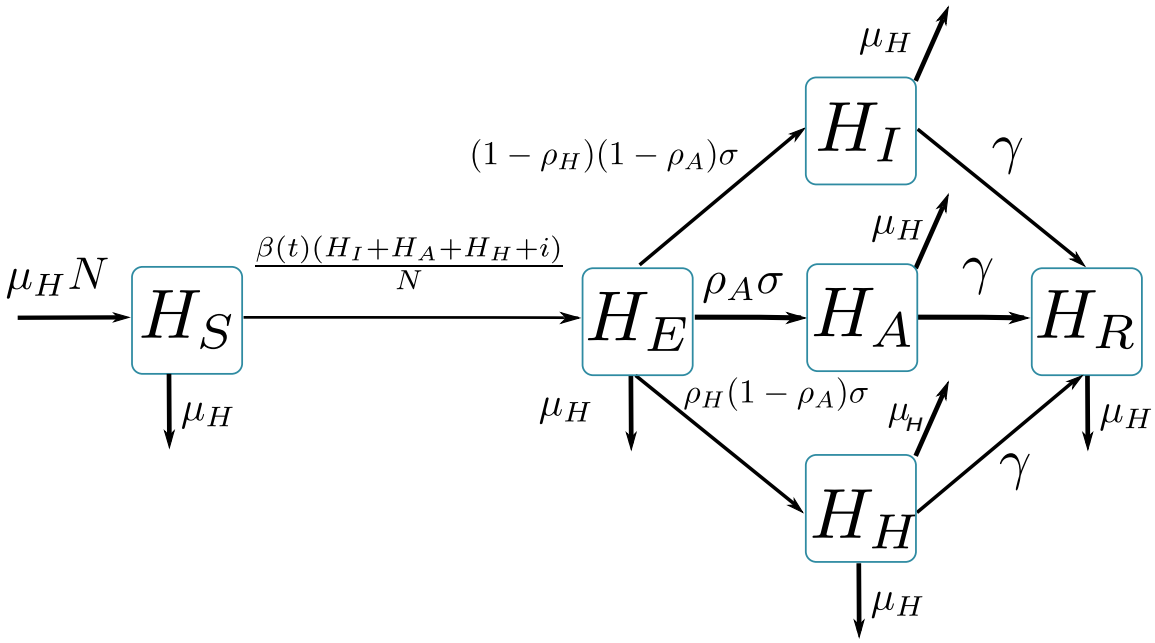

Figure 5 – **Graphical representation of SEIAR model.** H_S susceptible individuals; H_E infected (not yet infectious) individuals; H_A asymptomatic infectious individuals; H_I mildly symptomatic infectious individuals; H_H hospitalized infectious individuals; H_R recovered individuals; $\beta(t)$ is the transmission parameter; σ is the rate at which H_E-individuals move to the infectious classes H_I, H_A and H_H; a proportion ρ_A of H_E-individuals do not show symptoms during the infectious period; a proportion ρ_H of symptomatic individuals go to hospital; infectious individuals (H_I,H_A,H_H) then recover at rate γ; individuals leave the children population at rate μ_H. $H_S+H_E+H_A+H_I+H_H+H_R = N$.

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

2.2.3 Model with two virus serotypes

In the 2012 and 2013 epidemics, DENV-1 was highly dominant : the three other serotypes represented less than 1% of the cases reported in the DENFREE study in 2012 and about 20% 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. Figure 6). 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 equation 5, called SEIR2 model). In this context, the reproduction numbers for each strain are equal, $R_0^{SEIR2_1}(t) = R_0^{SEIR2_2}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$.

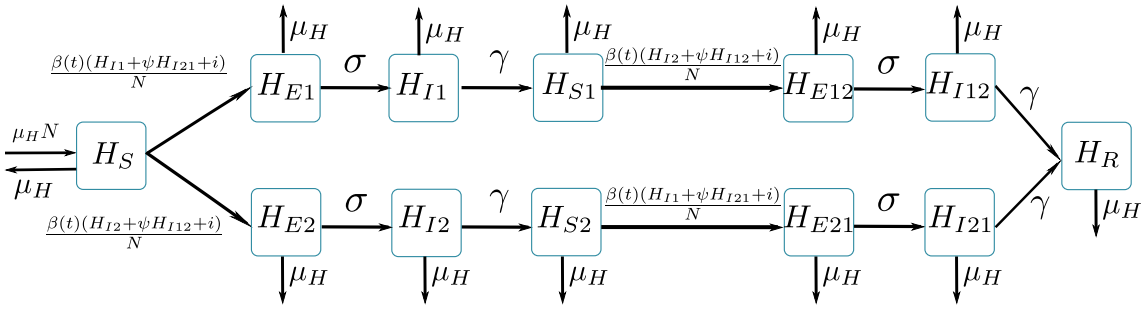

Figure 6 – **Graphical representation of SEIR2 models.** H_S individuals susceptible to both strains; H_{E1} (resp. H_{E2}) individuals infected (not yet infectious) to strain 1 (resp. strain 2); H_{I1} (resp. H_{I2}) individuals infectious to strain 1 (resp. strain 2); H_{S1} (resp. H_{S2}) individuals 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; σ is the rate at which exposed individuals move to the infectious class; infectious individuals then recover at rate γ; ψ is the change in infectivity for secondary infected individuals in SEIR2psi model (in SEIR2 model, $\psi = 1$); individuals leave the children population at rate μ_H. $H_S + H_{E1} + H_{E2} + H_{I1} + H_{I2} + H_{S1} + H_{S2} + H_{E12} + H_{E21} + H_{I12} + H_{I21} + H_R = N$.

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

We also considered another version of the model including interaction between strains, 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 [7]. In our model (called SEIRpsi model), primary and secondary infections differ in infectiousness, through a parameter ψ [35]. This parameter is estimated between zero and three: values superior to 1 correspond to transmission cross-enhancement (because of higher virus titers during secondary infections [35]) 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 [36]). As in Ferguson et al. [35], we define $R_0^{SEIR2_1}(t) = R_0^{SEIR2_2}(t) = \frac{\beta(t)\sigma}{(\gamma+\mu_H)(\sigma+\mu_H)}$ the basic reproduction number for each strain.

In the DENFREE data, DENV-1 cases are assumed to be either H_{I1} or H_{I21} and DENV-2/ DENV-3/ DENV-4 cases H_{I2} or H_{I12}. In the NDSS data, observed cases are assumed to be H_{I1} , H_{I2}, H_{I12} or H_{I21}.

2.2.4 Seasonality

All models include seasonality through the use of a time-varying transmission parameter $\beta(t) = \beta[1 + b.sin(2\pi(\frac{t}{365} + p)]$, 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.

Parameter		Prior distribution	Reference	Models
Infectiousness, incubation and mortality rates				
γ^{-1}	infectious period (days)	4.5	[3]	All
σ^{-1}	intrinsic incubation period (days)	5.9	[37]	Pandey, Laneri
τ^{-1}	extrinsic incubation period (days)	10	[37]	Pandey, Laneri
σ^{-1}	both incubation periods (days)	15.9	[37]	SEIR, SEIAR, SEIR2, SEIR2psi
μ_H^{-1}	age duration (years)	15	assumed	All
μ_V^{-1}	mosquito lifespan (days)	15	[38]	Pandey
Transmission parameters				
R_0	average basic reproduction number	Uniform[0, 20]	assumed	All
β_V	transmission from human to mosquito	Uniform[0.1, 2]	[29]	Pandey
ψ	inh./enh. of infectiousness	Uniform[0, 3]	[39]	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]	[40]	All
$v_I(0)$ or $\lambda(0)$	Initial number of infected mosquitoes	Uniform[10^{-6},10^{-3}]	assumed	Pandey or Laneri
$v_E(0)$ or $\kappa(0)$	Initial number of exposed mosquitoes	$v_I(0)$ or $\lambda(0)$	assumed	Pandey or Laneri
$H_H(0)$, $H_A(0)$	Initial number of exposed individuals	$H_I(0)$	assumed	SEIAR
$H_{S1}(0)$, $H_{S2}(0)$	Initial number of exposed individuals	N*Uniform[0.1, 1]	assumed	SEIR2, SEIR2psi
$H_{I1}(0)$, $H_{I2}(0)$	Initial number of exposed individuals	Uniform[0, 100]	assumed	SEIR2, SEIR2psi
$H_{E1}(0)$, $H_{I21}(0)$, $H_{E21}(0)$	Initial number of exposed individuals	$H_{I1}(0)$	assumed	SEIR2, SEIR2psi
$H_{E2}(0)$, $H_{I12}(0)$, $H_{E12}(0)$	Initial number of exposed individuals	$H_{I2}(0)$	assumed	SEIR2, SEIR2psi
Observation process				
r_N	observation rate for NDSS data	Uniform[0, 1]	assumed	All
r_D	observation rate for DENFREE data	fixed	assumed	All
ρ_A	Proportion of asymptomatic cases	Uniform[0, 1]	assumed	All
ρ_H	Proportion of hospitalized cases	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

Table 3 – **Prior distributions of parameters**. "Uniform[0,20]" indicates a uniform distribution in the range [0,20]. "Normal(4.5,0.1) in [4,5]" indicates a normal distribution with mean 4.5 and standard deviation 0.1, restricted to the range [4,5].

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 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 β (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 [40], 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_λ 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_λ is used as the mean of the gaussian prior, and the standard deviation is fixed at 0.05.

2.3.2 Observation rate

Using information on the sampling scheme, we calculate an observation rate on DENFREE data (cf. Table 4). 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.

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_s=(g/f)*(n/N) and r_a=b/N)	**0.0283**	**0.0107**	**0.0661**	**0.0255**

Table 4 – **Calculation of the observation rate for DENFREE data.** These calculations include all cases under 15 years old in the DENFREE study (18th Jun - 24th Dec 2012 and 3rd Jun - 23rd Sep 2013).

Different observation rates are used in the SEIAR model. For DENFREE cases, the observation rate is assumed known and relies on the fact that the entire population was not tested during the study. For NDSS data, the observation rate is interpreted as the proportion of hospitalized cases (which assumes that all hospitalized cases go through surveillance and neglects the presence of private hospitals or non reports from hospitals).

2.4 Estimation

Models are performed in deterministic framework and estimations are made using random walk Metropolis Hastings. SSM software [41] is used for simulations and calculations. The observation model is negative binomial with the dispersion parameter fixed at 0.1 [42]. In order to initialize the starting point of MCMC chains, we ran a simplex algorithm on 10,000 parameter sets sampled with latin hypercube sampling (with *lhs* R package [43]) and chose the one with the highest posterior value at the end of the chain. The posterior is highly multimodal and thanks to this initialization, the MCMC chain explored the region of the highest mode. The covariance matrix of the proposal distribution was initialized using adaptative MCMC as in Dureau et al. [41].

2.5 Model comparison

2.5.1 Statistical indicators

In order to identify the best model, the Deviance Information Criterion (DIC) [44] 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 mean RMSE (root mean square error) between observations and simulations of the model with parameters sampled in the posterior distribution (with observation indices from t_0 to T, and N simulated trajectories):

$$\text{mean RMSE} = \frac{1}{N}\sum_{i=1}^{N}\text{RMSE}_i = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{T-t_0}\sum_{t=t_0}^{T}(y_{i,t}-y_t^{obs})^2}$$

We first calculated this indicator on the data used for estimation (separating NDSS data and DENFREE data). Then we used it to assess the predictive performance of the model, comparing projections of the model with NDSS observations for 2014 and 2015.

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 infection (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 infection (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).

	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, $i=1,2$)	$R_0(t)\frac{H_S(t)}{N}$	$R_0(t)\frac{H_S(t)}{N}v_s(t)$	$R_0(t)\frac{H_S(t)+H_{Sj}(t)}{N}$	$R_0(t)\frac{H_S(t)+H_{Sj}(t)}{N}\left[\frac{H_{Ii}(t)+\psi H_{Iji}(t)}{H_{Ii}(t)+H_{Iji}(t)}\right]$

Table 5 – **Reproduction numbers calculation in each model.**

Calculations are made using R version 3.2.2 [45], and graphics using *ggplot2* [46].

3 Results

3.1 Statistical comparison

Model	**SEIR**	**Laneri**	**Pandey**	**SEIAR**	**SEIR2**	**SEIR2psi**
nb parameters	7	8	9	8	10	11
nb observations	656	656	656	686	686	686
ESTIMATION SET						
DIC	4774	4818	4778	4990	4800	**4604**
mean RMSE NDSS	23.6	23.8	23.8	24.6	18.8	**16.8**
mean RMSE NDSS 2002	*52.2*	*51.4*	*52.6*	*55.9*	*14.4*	*15.1*
mean RMSE NDSS 2003-2013	*19.0*	*19.4*	*19.2*	*19.4*	*19.1*	*17.0*
mean RMSE DENFREE	23.7	23.8	23.0	24.3	**20.3**	20.7
TEST SET						
mean RMSE 2014-2015	14.7	14.7	15.3	15.4	21.3	**9.8**
mean RMSE 2014	*18.5*	*18.5*	*18.9*	*19.5*	*25.8*	*11.4*
mean RMSE 2015	*9.6*	*9.1*	*10.2*	*9.6*	*15.4*	*7.7*

Table 6 – **Information criteria in deterministic model.** DIC is the Deviance Information Criterion [44]. RMSE is the root mean square error 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, $\frac{1}{12}(\text{RMSE}_i^{\text{NDSS 2002}})^2 + \frac{11}{12}(\text{RMSE}_i^{\text{NDSS 2003-2013}})^2 = (\text{RMSE}_i^{\text{NDSS}})^2$. Convergence diagnosis are displayed in Appendix (trace plots et correlation plots respectively in Figures 12 and 13).

For single strain models, the SEIR and Pandey models 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 and 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. Figures 7-8), with a period of approximately six years, and a large overestimation of the 2002 outbreak (the RMSE for 2002 is far higher (> 50) than the average for the other years (<20)). Due to the small number of observed asymptomatic cases, SEIAR model also produces a similar dynamic. Models with two strains outperform all the other models in terms of RMSE, but the difference is mainly explained by the first year of simulation (2002). When visualizing the simulations compared to the data (cf. Figures 7-8), all models underestimate a large number of epidemic peaks and all models but the ones with two 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.

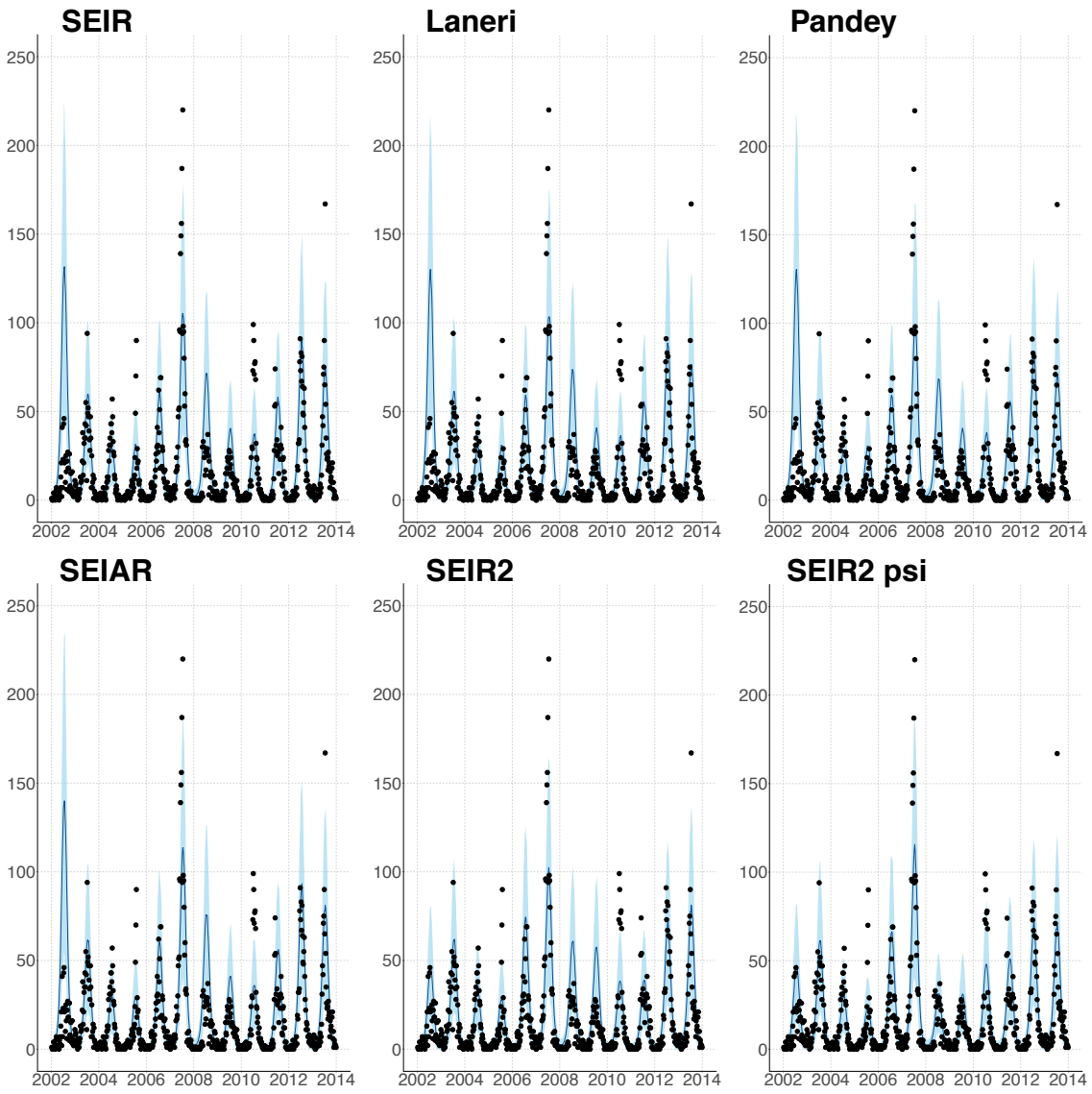

Figure 7 – **Number of observed cases per week and NDSS data, 2002-2013.** Simulations with negative binomial noise using parameters from the posterior distribution (SEIR/Laneri/Pandey/SEIR2/SEIR2psi: observed NDSS cases, SEIAR: hospitalized cases). Posterior median (solid line), 95% credible intervals (shaded blue area) and NDSS data points (black dots).

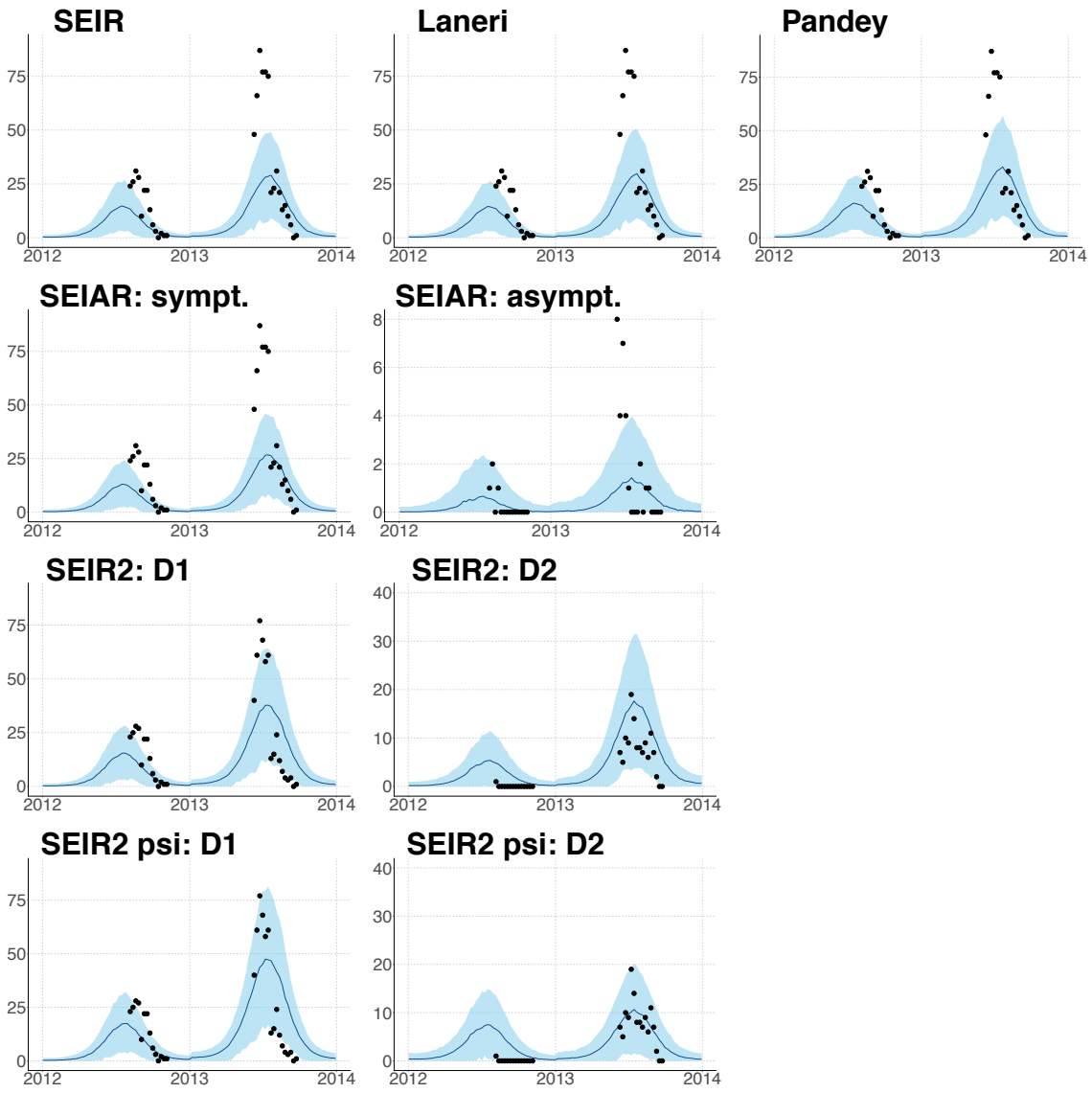

Figure 8 – **Number of observed cases per week and DENFREE data, 2012-2013.** Simulations with negative binomial noise using parameters from the posterior distribution. Posterior median (solid line), 95% credible intervals (shaded blue area) and DENFREE data points (black dots).

Considering the predictive capacity of the models, only SEIR2psi model predicts a smaller epidemic in 2014, as was observed in the data (cf. Figure 9). Across the other models, the predictions for 2014-2015 are qualitatively similar: they largely overestimate the 2014 epidemic, and to a lesser extent the one in 2015.

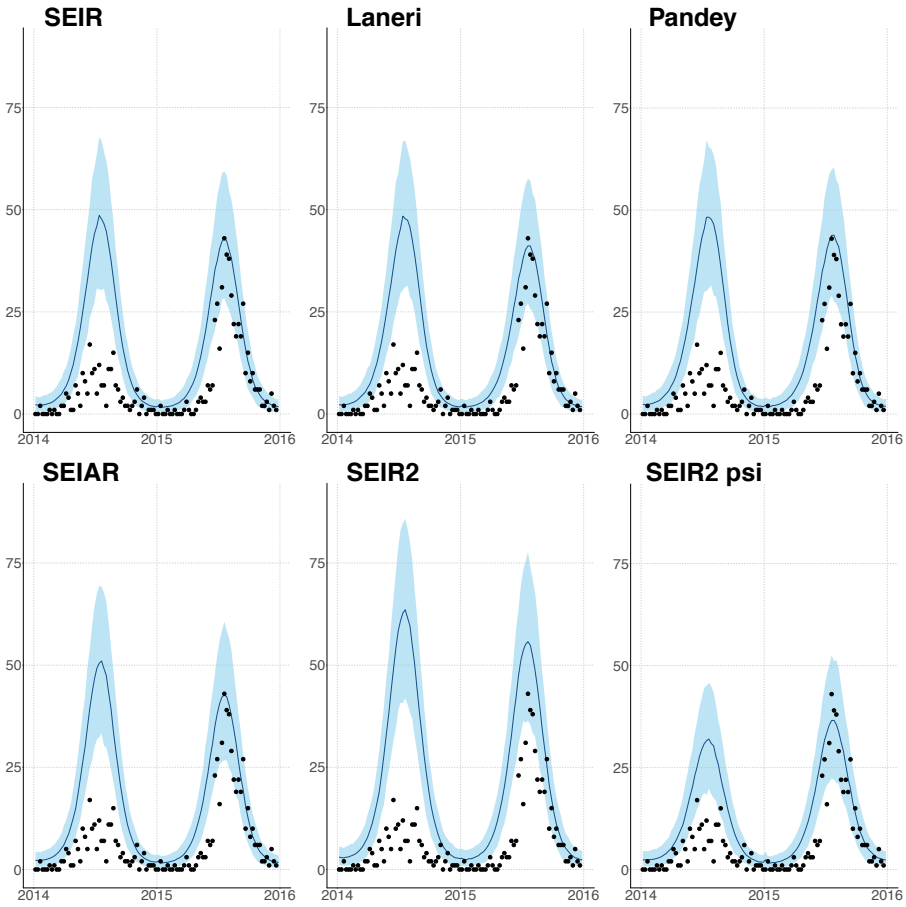

Figure 9 – **Projections of the number of observed cases per week and NDSS data, 2014-2015.** Simulations with negative binomial noise using parameters from the posterior distribution (SEIR/Laneri/Pandey/SEIR2/SEIR2psi: observed NDSS cases, SEIAR: hospitalized cases). Posterior median (solid line), 95% credible intervals (shaded blue area) and NDSS data points (black dots).

3.2 Epidemiological comparison

Model		**SEIR**	**Laneri**	**Pandey**	**SEIAR**	**SEIR2**	**SEIR2psi**
mean R_0	median (95%CI)	2.31 (2.27-2.36)	2.38 (2.33-2.42)	3.35 (3.27-3.43)	2.30 (2.26-2.34)	2.06 (2.02-2.11)	2.53 (2.47-2.60)
max R_0	median (95%CI)	3.72 (3.64-3.81)	3.75 (3.67-3.83)	6.01 (5.90-6.24)	3.72 (3.64-3.80)	3.24 (3.17-3.32)	3.93 (3.83-4.03)
ψ	median (95%CI)	.	.	.	.	.	0.69 (0.66-0.73)
$H_S(0)$**/N (%)**	median (95%CI)	47 (46-47)	47 (46-48)	37 (36-37)	47 (47-48)	33 (29-35)	27 (24-29)
$H_{S1}(0)$**/N (%)**	median (95%CI)	.	.	.	.	20 (18-27)	22 (20-24)
$H_{S2}(0)$**/N (%)**	median (95%CI)	.	.	.	.	11 (10-18)	12 (10-15)
Observation rate (%)	median (95%CI)	17 (16-18)	17 (17-18)	14 (14-15)	.	10 (9-10)	8 (8-8)
Hospitalized (%)	median (95%CI)	.	.	.	20 (19-21)	.	
Asymptomatic (%)	median (95%CI)	.	.	.	12 (10-14)	.	
Median annual incidence proportion							
primary infections (%)	median 2002-2015 (min - max)	8 (4-16)	7 (4-16)	12 (7-25)	7 (4-17)	13 (7-21)	14 (8-31)
***secondary infections* (%)**	median 2002-2015 (min - max)	.	.	.	.	7 (3-11)	7 (3-17)
mean R_e	median (95%CI)	1.03 (1.03-1.04)	1.08 (1.07-1.08)	1.16 (1.15-1.17)	1.04 (1.03-1.04)	.	.
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)	.	.	.	.	1.03 (1.03-1.03)	1.03 (1.02-1.03)
max R_e	median (95%CI)	1.76 (1.73-1.79)	1.79 (1.76-1.82)	2.27 (2.22-2.34)	1.77 (1.75-1.80)	.	.
max* R_e *strain 1	median (95%CI)	.	.	.	.	1.78 (1.76-1.81)	1.80 (1.77-1.82)
max* R_e *strain 2	median (95%CI)	.	.	.	.	1.73 (1.70-1.75)	1.71 (1.68-1.74)

Table 7 – **Epidemiological criteria among children under 15 years old.** Estimated parameters from the posterior distribution and indicators based on simulations over 2002-2015.

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 6.0). The estimated values are very close in the SEIR, Laneri and SEIAR models. The estimates are on the lower side of those estimates for South East Asia in general [47, 48, 49]. In particular, our estimates are close to the estimates for R_0 in Cambodia made from age-stratified case-notification data [49], as far as the estimates based on different data types and models are comparable. 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. The effective reproductive numbers (R_e) have a mean value around 1 and a maximum value

around 1.8 in most models, but it is higher in the Pandey model.

In SEIR2psi model, the parameter ψ quantifying the interaction between strains is inferior to 1, suggesting a reduced infectivity of secondary infections on average, as in Aguiar et al. [36], or in Coudeville and Garnett [50]. This suggests that at the population level cross-protection 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. This proportion is lower in the Pandey model. In the models with two strains, the number of susceptibles to both strains is smaller, but as a whole, less children are immune to the disease, since more than 30% have experienced only one infection. The proportion of children who are susceptible to one or both strains are however correlated in the MCMC chain (cf. Figure 13 in Appendix), indicating that their relative shares are not well identified. These values are in the range of the measures of seroprevalence in several Asian countries [51]. As the measures reveal large differences between countries [51, 40, 52, 53], 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 14 to 21% in models with one strain and between 6 and 13% in the two strain models. 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. Figure 10). 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 [13] or Singapore [54].

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. Figure 10 for SEIR2psi and SEIR models), allowing the possibility for large outbreaks to occur in the future.

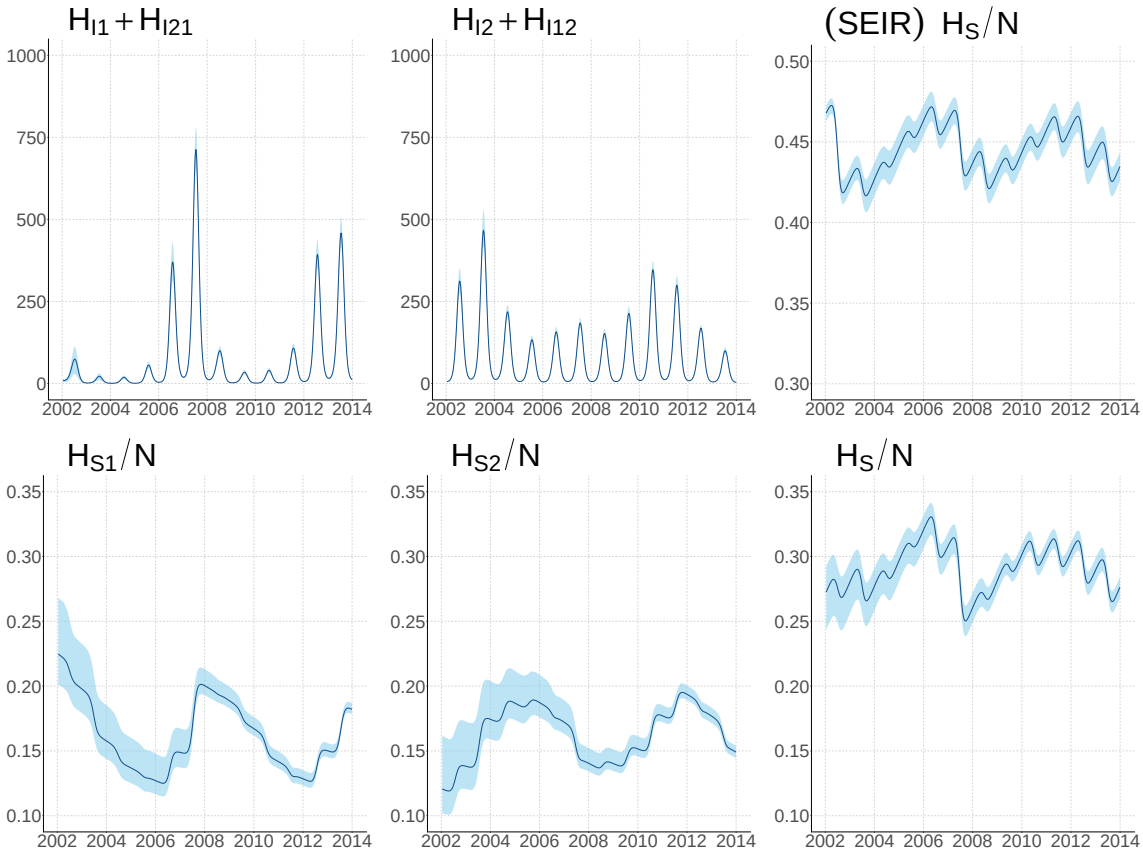

Figure 10 – **Current number of infected and proportion of susceptible individuals per serotype in SEIR2psi model.** Median of the simulations and confidence intervals without observation error. For comparison, the proportion of susceptibles in SEIR model is also displayed.

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. Figure 11). The values for primary infections are coherent with other studies in Vietnam or Indonesia who analyzed

seroprevalence data or seroconversion data [40, 55, 56, 57, 52]. The incidence proportion is highly variable from one year to the next, especially in models with two strains.

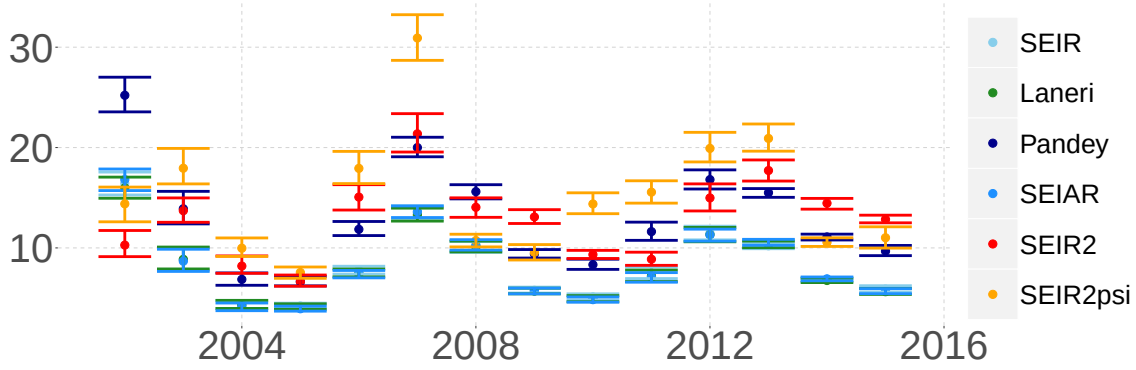

Figure 11 – **Annual incidence proportion of first dengue infection (%).** Median and confidence intervals per year, based on simulations over 2002-2015.

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 their epidemiological features. In particular, in our models, the proportion of susceptibles displays year-to-year variations but remains high over the period, so that large outbreaks may occur in the future. 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 [7], 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 "house-to-house movement" [58, 59]. This feature was previously analytically studied and coherent with dengue incidence time series in Thailand [36, 60]. On the contrary, including vectorial transmission or a compartment for asymptomatic infections did not seem to improve the model fit despite the additional complexity. The non-utility of including vectorial components has been observed by several

authors previously [29, 61, 62]. 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. 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 [63, 61, 62]. The main effect of mosquito dynamics is captured in our model by the seasonal forcing. The lack of additional improvement when including 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. Yoon et al. [59] also observe many inapparent cases in their cohort but few strictly asymptomatic cases in their cluster 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.

We also obtained some insight on the parameters describing transmission. The average annual R_0 is estimated between 2.1 and 2.6 in most 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 the uncertainties due to diverse estimation procedures. The median annual incidence at primary infection over the period is between 8 and 14, with large year-to-year variations. The estimated observation rate on surveillance data varies between models, (14-21% in models with one strain and 6-13% in two strain models), indicating in both cases a high proportion of unreported infections. These values are in line with the large underrecognition highlighted in Wichmann et al. [64], and also with other studies in South-East Asia [59, 55, 57].

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 R_0 or incidence.

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 [39]. 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 [12, 65]. 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, maybe due to particular climatic conditions. In many countries in South-East Asia, except Malaysia, the reported incidence was lower than in 2013 [66]. Many provinces of Thailand also reported fewer cases than usual in 2014 [67]. 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 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. As in our previous work [34], 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.

In conclusion, our analyses highlight the importance of using two-strain models with interacting effects and the lack of added value of incorporating vector and asymptomatic components. 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 very low incidence of asymptomatic infections questions their incorporation explicitly in the models. Another important aspect is related to choosing the best model considering the available data. In quantitative epidemiology, in recent years, it has been standard to use statistical indices computed on estimated likelihood [29, 28]. In this work, we discussed not only the goodness of fit of the models, but also their prediction capacity and epidemiological features. The sole use of statistical indices is not enough and it is crucial to take into account some epidemiological features (such as R_0, annual incidence, dynamics of the susceptible class) of the disease studied into account.

Acknowledgments

All the authors have been supported by a grant from Agence Nationale de la Recherche for the PANIC project (Targeting PAthogen's NICche: 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.

References

[1] Grange L, Simon-Loriere E, Sakuntabhai A, Gresh L, Paul R, Harris E. Epidemiological Risk Factors Associated with High Global Frequency of Inapparent Dengue Virus Infections. Frontiers in Immunology. 2014 Jun;5:280. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4052743/.

[2] Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, et al. Dengue: a continuing global threat. Nature Reviews Microbiology. 2010 Dec;8:S7–16.

[3] World Health Organisation. Dengue;. Accessed: 2017-07-20. Available from: http://www.who.int/topics/dengue/en/.

[4] Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, et al. Global spread of dengue virus types: mapping the 70 year history. Trends in Microbiology. 2014 Mar;22(3):138–146.

[5] Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496(7446):504–507. Available from: http://www.nature.com/nature/journal/v496/n7446/full/nature12060.html.

[6] Shepard DS, Undurraga EA, Halasa YA, Stanaway JD. The global economic burden of dengue: a systematic analysis. The Lancet Infectious Diseases. 2016 Aug;16(8):935–941.

[7] Halstead SB. Dengue. The Lancet. 2007 Nov;370(9599):1644–1652. Available from: http://www.sciencedirect.com/science/article/pii/S0140673607616870.

[8] Messina JP, Brady OJ, Pigott DM, Golding N, Kraemer MUG, Scott TW, et al. The many projected futures of dengue. Nature Reviews Microbiology. 2015 Apr;13(4):230–239.

[9] Sharp TM, Tomashek KM, Read JS, Margolis HS, Waterman SH. A New Look at an Old Disease: Recent Insights into the Global Epidemiology of Dengue. Current Epidemiology Reports. 2017;4(1):11–21.

[10] Reiner RC, Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. Journal of The Royal Society Interface. 2013 Feb;10(81):20120921–20120921. Available from: http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2012.0921.

[11] Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. Science. 2015 Mar;347(6227):aaa4339–aaa4339. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.aaa4339.

[12] Salje H, Lessler J, Endy TP, Curriero FC, Gibbons RV, Nisalak A, et al. Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. Proceedings of the National Academy of Sciences of the United States of America. 2012 Jun;109(24):9535–9538.

[13] Reich NG, Shrestha S, King AA, Rohani P, Lessler J, Kalayanarooj S, et al. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity.

Journal of The Royal Society Interface. 2013 Jul;10(86):20130414–20130414. Available from: http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2013.0414.

[14] Clapham H, Cummings DAT, Nisalak A, Kalayanarooj S, Thaisomboonsuk B, Klungthong C, et al. Epidemiology of Infant Dengue Cases Illuminates Serotype-Specificity in the Interaction between Immunity and Disease, and Changes in Transmission Dynamics. PLOS Neglected Tropical Diseases. 2015 Dec;9(12):e0004262. Available from: http://dx.plos.org/10.1371/journal.pntd.0004262.

[15] Nisalak A, Clapham HE, Kalayanarooj S, Klungthong C, Thaisomboonsuk B, Fernandez S, et al. Forty Years of Dengue Surveillance at a Tertiary Pediatric Hospital in Bangkok, Thailand, 1973-2012. The American Journal of Tropical Medicine and Hygiene. 2016 Jun;94(6):1342–1347.

[16] Strickman D, Sithiprasasna R, Kittayapong P, Innis BL. Distribution of dengue and Japanese encephalitis among children in rural and suburban Thai villages. The American Journal of Tropical Medicine and Hygiene. 2000 Aug;63(1-2):27–35.

[17] Mammen Jr MP, Pimgate C, Koenraadt CJ, Rothman AL, Aldstadt J, Nisalak A, et al. Spatial and temporal clustering of dengue virus transmission in Thai villages. PLoS Med. 2008;5(11):e205. Available from: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0050205.

[18] Aldstadt J, Yoon IK, Tannitisupawong D, Jarman RG, Thomas SJ, Gibbons RV, et al. Space-time analysis of hospitalised dengue patients in rural Thailand reveals important temporal intervals in the pattern of dengue virus transmission. Tropical medicine & international health: TM & IH. 2012 Sep;17(9):1076–1085.

[19] Guha-Sapir D, Schimmer B. Dengue fever: new paradigms for a changing epidemiology. Emerging Themes in Epidemiology. 2005;2:1. Available from: http://dx.doi.org/10.1186/1742-7622-2-1.

[20] Muhammad Azami NA, Salleh SA, Neoh Hm, Syed Zakaria SZ, Jamal R. Dengue epidemic in Malaysia: Not a predominantly urban disease anymore. BMC Research Notes. 2011;4:216. Available from: http://dx.doi.org/10.1186/1756-0500-4-216.

[21] Chareonsook O, Foy HM, Teeraratkul A, Silarug N. Changing epidemiology of dengue hemorrhagic fever in Thailand. Epidemiology and Infection. 1999 Feb;122(1):161–166.

[22] Vong S, Khieu V, Glass O, Ly S, Duong V, Huy R, et al. Dengue Incidence in Urban and Rural Cambodia: Results from Population-Based Active Fever Surveillance, 2006–2008. PLoS Neglected Tropical Diseases. 2010 Nov;4(11):e903. Available from: http://dx.plos.org/10.1371/journal.pntd.0000903.

[23] Reller ME, Bodinayake C, Nagahawatte A, Devasiri V, Kodikara-Arachichi W, Strouse JJ, et al. Unsuspected Dengue and Acute Febrile Illness in Rural and Semi-Urban Southern Sri Lanka. Emerging Infectious Diseases. 2012 Feb;18(2):256–263. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3310451/.

[24] Teurlai M, Huy R, Cazelles B, Duboz R, Baehr C, Vong S. Can Human Movements Explain Heterogeneous Propagation of Dengue Fever in Cambodia? PLoS Neglected Tropical Diseases. 2012 Dec;6(12):e1957. Available from: http://dx.plos.org/10.1371/journal.pntd.0001957.

[25] Cuong HQ, Vu NT, Cazelles B, Boni MF, Thai KTD, Rabaa MA, et al. Spatiotemporal Dynamics of Dengue Epidemics, Southern Vietnam. Emerging Infectious Diseases. 2013 Jun;19(6):945–953. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3713821/.

[26] Cazelles B, Cazelles K. Major urban centers have weak influence on the timing of dengue epidemics in Southeast Asia. International Journal of Infectious Diseases.

2014 Apr;21, Supplement 1:217. Available from: http://www.sciencedirect.com/science/article/pii/S1201971214009321.

[27] King AA, Ionides EL, Pascual M, Bouma MJ. Inapparent infections and cholera dynamics. Nature. 2008 Aug;454(7206):877–880. Available from: http://www.nature.com/doifinder/10.1038/nature07084.

[28] Camacho A, Ballesteros S, Graham AL, Carrat F, Ratmann O, Cazelles B. Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. Proceedings of the Royal Society of London B: Biological Sciences. 2011 Apr;p. rspb20110300. Available from: http://rspb.royalsocietypublishing.org/content/early/2011/04/22/rspb.2011.0300.

[29] Pandey A, Mubayi A, Medlock J. Comparing vector–host and SIR models for dengue transmission. Mathematical Biosciences. 2013 Dec;246(2):252–259. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0025556413002435.

[30] Duong V, Lambrechts L, Paul RE, Ly S, Lay RS, Long KC, et al. Asymptomatic humans transmit dengue virus to mosquitoes. Proceedings of the National Academy of Sciences. 2015 Nov;Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1508114112.

[31] Huy R, Buchy P, Conan A, Ngan C, Ong S, Ali R, et al. National dengue surveillance in Cambodia 1980–2008: epidemiological and virological trends and the impact of vector control. Bulletin of the World Health Organization. 2010 Sep;88(9):650–657. Available from: http://www.who.int/bulletin/volumes/88/9/09-073908.pdf.

[32] National Institute of Statistics, Ministry of Planning, Phnom Penh, Cambodia. General Population Census of Cambodia 2008. National report on final census results; 2009.

[33] Laneri K, Bhadra A, Ionides EL, Bouma M, Dhiman RC, Yadav RS, et al. Forcing Versus Feedback: Epidemic Malaria and Monsoon Rains in Northwest India. PLoS

Computational Biology. 2010 Sep;6(9):e1000898. Available from: http://dx.plos.org/10.1371/journal.pcbi.1000898.

[34] Champagne C, Salthouse DG, Paul R, Cao-Lormeau VM, Roche B, Cazelles B. Structure in the variability of the basic reproductive number (R0) for Zika epidemics in the Pacific islands. eLife. 2016 Nov;5:e19874. Available from: https://elifesciences.org/content/5/e19874v2.

[35] Ferguson N, Anderson R, Gupta S. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. Proceedings of the National Academy of Sciences of the United States of America. 1999 Jan;96(2):790–794. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC15215/.

[36] Aguiar M, Ballesteros S, Kooi BW, Stollenwerk N. 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. Journal of theoretical biology. 2011;289:181–196. Available from: http://www.sciencedirect.com/science/article/pii/S0022519311004462.

[37] Chan M, Johansson MA. The Incubation Periods of Dengue Viruses. PLoS ONE. 2012 Nov;7(11):e50972. Available from: http://dx.doi.org/10.1371/journal.pone.0050972.

[38] Liu-Helmersson J, Stenlund H, Wilder-Smith A, Rocklöv J. Vectorial Capacity of Aedes aegypti: Effects of Temperature and Implications for Global Dengue Epidemic Potential. PLoS ONE. 2014 Mar;9(3):e89783. Available from: http://dx.plos.org/10.1371/journal.pone.0089783.

[39] Aguiar M, Kooi BW, Rocha F, Ghaffari P, Stollenwerk N. How much complexity is needed to describe the fluctuations observed in dengue hemorrhagic fever incidence

data? Ecological Complexity. 2013 Dec;16:31–40. Available from: http://www.sciencedirect.com/science/article/pii/S1476945X12000670.

[40] Thai KT, Binh TQ, Giao PT, Phuong HL, Hung LQ, Nam NV, et al. Seroprevalence of dengue antibodies, annual incidence and risk factors among children in southern Vietnam. Tropical Medicine & International Health. 2005;10(4):379–386. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3156.2005.01388.x/full.

[41] Dureau J, Ballesteros S, Bogich T. SSM: Inference for time series analysis with State Space Models; 2013. Available from: https://github.com/JDureau/ssm/blob/master/doc/doc.pdf.

[42] Bretó C, He D, Ionides EL, King AA. Time series analysis via mechanistic models. The Annals of Applied Statistics. 2009 Mar;3(1):319–348. Available from: http://projecteuclid.org/euclid.aoas/1239888373.

[43] Carnell R. Package 'lhs'; 2016. Available from: http://cran.stat.auckland.ac.nz/web/packages/lhs/lhs.pdf.

[44] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002 Oct;64(4):583–639. Available from: http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00353/abstract.

[45] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available from: http://www.R-project.org.

[46] Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2009. Available from: http://ggplot2.org.

[47] Johansson MA, Hombach J, Cummings DAT. Models of the impact of dengue vaccines: a review of current research and potential approaches. Vaccine. 2011 Aug;29(35):5860–5868. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4327892/.

[48] Imai N, Dorigatti I, Cauchemez S, Ferguson NM. Estimating Dengue Transmission Intensity from Sero-Prevalence Surveys in Multiple Countries. PLoS Neglected Tropical Diseases. 2015 Apr;9(4). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400108/.

[49] Imai N, Dorigatti I, Cauchemez S, Ferguson NM. Estimating Dengue Transmission Intensity from Case-Notification Data from Multiple Countries. PLOS Negl Trop Dis. 2016;10(7):e0004833. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004833.

[50] Coudeville L, Garnett GP. Transmission Dynamics of the Four Dengue Serotypes in Southern Vietnam and the Potential Impact of Vaccination. PLOS ONE. 2012 12;7(12):1–11. Available from: https://doi.org/10.1371/journal.pone.0051244.

[51] L'Azou M, Moureau A, Sarti E, Nealon J, Zambrano B, Wartel TA, et al. Symptomatic dengue in children in 10 Asian and Latin American countries. New England Journal of Medicine. 2016;374(12):1155–1166. Available from: http://dx.doi.org/10.1056/NEJMoa1503877.

[52] Prayitno A, Taurel AF, Nealon J, Satari HI, Karyanti MR, Sekartini R, et al. Dengue seroprevalence and force of primary infection in a representative population of urban dwelling Indonesian children. PLOS Neglected Tropical Diseases. 2017;11(6):e0005621. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0005621.

[53] Vongpunsawad S, Intharasongkroh D, Thongmee T, Poovorawan Y. Seroprevalence of antibodies to dengue and chikungunya viruses in Thailand. PLOS ONE. 2017;12(6):e0180560. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0180560.

[54] Ministry of Health, Singapore. Dengue review of 2016 and outlook for 2017. Epidemiol News Bull. 2017 Apr;43(2):47–56. Available from: https://www.moh.gov.sg/content/dam/moh_web/Publications/Epidemiological%20News%20Bulletin/ENB%20Quarterly_Apr%202017%20Vol%2043%20No%202%20(Final).pdf.

[55] Thai KTD, Nga TTT, Van Nam N, Phuong HL, Giao PT, Hung LQ, et al. Incidence of primary dengue virus infections in Southern Vietnamese children and reactivity against other flaviviruses. Tropical Medicine & International Health. 2007 Dec;12(12):1553–1557. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3156.2007.01964.x/abstract.

[56] Tien NTK, Luxemburger C, Toan NT, Pollissard-Gadroy L, Huong VTQ, Van Be P, et al. A prospective cohort study of dengue infection in schoolchildren in Long Xuyen, Viet Nam. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2010 Sep;104(9):592–600. Available from: https://academic.oup.com/trstmh/article-lookup/doi/10.1016/j.trstmh.2010.06.003.

[57] Graham RR, Juffrie M, Tan R, Hayes CG, Laksono I, Ma'roef C, et al. A prospective seroepidemiologic study on dengue in children four to nine years of age in Yogyakarta, Indonesia I. studies in 1995-1996. The American Journal of Tropical Medicine and Hygiene. 1999 Sep;61(3):412–419.

[58] Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan VA, Vazquez-Prokopec GM, Astete H, et al. House-to-house human movement drives dengue virus transmission.

652 Proceedings of the National Academy of Sciences of the United States of America.
653 2013 jan;110(3). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3549073/.

654 [59] Yoon IK, Rothman AL, Tannitisupawong D, Srikiatkhachorn A, Jarman RG,
655 Aldstadt J, et al. Underrecognized Mildly Symptomatic Viremic Dengue Virus
656 Infections in Rural Thai Schools and Villages. The Journal of Infectious Diseases.
657 2012 Aug;206(3):389–398. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/
658 PMC3490697/.

659 [60] Aguiar M, Paul R, Sakuntabhai A, Stollenwerk N. Are we modelling the correct
660 dataset? Minimizing false predictions for dengue fever in Thailand. Epidemiology &
661 Infection. 2014 Nov;142(11):2447–2459. Available from: http://journals.cambridge.org/
662 article_S0950268813003348.

663 [61] Rocha F, Aguiar M, Souza M, Stollenwerk N. Time-scale separation and centre
664 manifold analysis describing vector-borne disease dynamics. International Journal
665 of Computer Mathematics. 2013 Oct;90(10):2105–2125. Available from: http://www.
666 tandfonline.com/doi/abs/10.1080/00207160.2013.783208.

667 [62] Rocha F, Mateus L, Skwara U, Aguiar M, Stollenwerk N. Understanding dengue
668 fever dynamics: a study of seasonality in vector-borne disease models. International
669 Journal of Computer Mathematics. 2016 Aug;93(8):1405–1422. Available from: http:
670 //dx.doi.org/10.1080/00207160.2015.1050961.

671 [63] Dye C, WIlliams BG. Nonlinearities in the Dynamics of Indirectly-Transmitted
672 Infections (or, does having a Vector makes a Difference ?). In: Grenfell BT, Dobson AP,
673 editors. Ecology of Infectious Diseases in Natural Populations. Cambridge University
674 Press; 1995. p. 260–279.

675 [64] Wichmann O, Yoon IK, Vong S, Limkittikul K, Gibbons RV, Mammen MP, et al.
676 Dengue in Thailand and Cambodia: An Assessment of the Degree of Underrecognized

Disease Burden Based on Reported Cases. PLoS Neglected Tropical Diseases. 2011 Mar;5(3):e996. Available from: http://dx.plos.org/10.1371/journal.pntd.0000996.

[65] Salje H, Lessler J, Berry IM, Melendrez MC, Endy T, Kalayanarooj S, et al. Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size. Science. 2017 Mar;355(6331):1302–1306. Available from: http://science.sciencemag.org/content/355/6331/1302.

[66] Cheng Q, Jing Q, Spear RC, Marshall JM, Yang Z, Gong P. The interplay of climate, intervention and imported cases as determinants of the 2014 dengue outbreak in Guangzhou. PLOS Neglected Tropical Diseases. 2017 06;11(6):1–24. Available from: https://doi.org/10.1371/journal.pntd.0005701.

[67] Reich NG, Lauer SA, Sakrejda K, Iamsirithaworn S, Hinjoy S, Suangtho P, et al. Challenges in Real-Time Prediction of Infectious Disease: A Case Study of Dengue in Thailand. PLOS Neglected Tropical Diseases. 2016;10(6):e0004761. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004761.

A Data and codes

The datasets and codes supporting this article are available at https://github.com/clchampag/KC-dengue.

B Convergence diagnosis

Figure 12 – **Trace plots for the MCMC algorithm, with 100,000 iterations.** a) SEIR. b) Laneri. c) Pandey. d) SEIAR. e) SEIR2. f) SEIR2psi

Figure 13 – **Correlation plots of the MCMC algorithm, with 100,000 iterations.** a) SEIR. b) Laneri. c) Pandey. d) SEIAR. e) SEIR2. f) SEIR2psi