

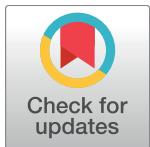
RESEARCH ARTICLE

Accounting for non-stationarity in epidemiology by embedding time-varying parameters in stochastic models

Bernard Cazelles^{1,2,3*}, Clara Champagne^{1,4}, Joseph Dureau⁵

1 Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, CNRS UMR 8197, Paris, France, **2** International Center for Mathematical and Computational Modeling of Complex Systems (UMMISCO), UMI 209, UPMC/IRD, France, **3** Hosts, Vectors and Infectious Agents, CNRS URA 3012, Institut Pasteur, Paris, France, **4** CREST, ENSAE, Université Paris Saclay, Palaiseau, France, **5** SNIPS, Paris, France

* cazelles@biologie.ens.fr



OPEN ACCESS

Citation: Cazelles B, Champagne C, Dureau J (2018) Accounting for non-stationarity in epidemiology by embedding time-varying parameters in stochastic models. PLoS Comput Biol 14(8): e1006211. <https://doi.org/10.1371/journal.pcbi.1006211>

Editor: Katia Koelle, Emory University, UNITED STATES

Received: December 16, 2017

Accepted: May 18, 2018

Published: August 15, 2018

Copyright: © 2018 Cazelles et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are in the Supporting Information in a zip file.

Funding: BC and CC have been supported by a grant from Agence Nationale de la Recherche for the PANIC project (ANR-14-CE02-0015-01). This research has also received funding from the European Commission Seventh Framework Program FP7 for the DENFREE project under Grant Agreement 282378. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The spread of disease through human populations is complex. The characteristics of disease propagation evolve with time, as a result of a multitude of environmental and anthropic factors, this non-stationarity is a key factor in this huge complexity. In the absence of appropriate external data sources, to correctly describe the disease propagation, we explore a flexible approach, based on stochastic models for the disease dynamics, and on diffusion processes for the parameter dynamics. Using such a diffusion process has the advantage of not requiring a specific mathematical function for the parameter dynamics. Coupled with particle MCMC, this approach allows us to reconstruct the time evolution of some key parameters (average transmission rate for instance). Thus, by capturing the time-varying nature of the different mechanisms involved in disease propagation, the epidemic can be described. Firstly we demonstrate the efficiency of this methodology on a toy model, where the parameters and the observation process are known. Applied then to real datasets, our methodology is able, based solely on simple stochastic models, to reconstruct complex epidemics, such as flu or dengue, over long time periods. Hence we demonstrate that time-varying parameters can improve the accuracy of model performances, and we suggest that our methodology can be used as a first step towards a better understanding of a complex epidemic, in situation where data is limited and/or uncertain.

Author summary

As our world becomes more and more globalized, infectious disease poses an ever-increasing threat to human health. The multitude of environmental and behavioral factors, which account for the spread of infectious diseases, are ever-evolving and thus infectious diseases propagation is complex.

In the face of this complexity, mathematical models offer valuable tools to study the dynamics of epidemic diseases. Developing adequate statistical and mathematical tools, that take account of the time-varying nature of the different mechanisms responsible for

Competing interests: The authors have declared that no competing interests exist.

disease propagation, remains a major challenge. To take this increasingly important aspect into consideration, we propose a flexible methodology that encompasses time-varying aspects of the epidemic. It does this via diffusion process equations for time-varying parameters. Considering the relative paucity of available data, our principal assertion is that it is preferable to use this flexible framework with time-varying parameters, that tracks epidemiological patterns, and updates the key parameters according to data, than to use a more complex model.

Introduction

Our world constantly faces the threat of emerging and re-emerging diseases and it has been shown that this has intensified over the past 50 years. This intensification is due, in part, to climate change, urbanization and globalization [1] meaning that infectious diseases remain a constant and unpredictable threat to human health.

Numerous factors contribute to the propagation of an infectious disease. These include increased human connectivity, limited availability of economic resources for adequate intervention, increasing antimicrobial resistance, evolution of the dominant strains and increasing parasite and vector resistance to the most widely used drugs and insecticides, etc. A key factor in this huge complexity is non-stationarity [2], meaning that the characteristics of the dynamical epidemiological processes evolve with time. Thus, the mechanisms of transmission are uncertain, making it difficult to obtain quantitative predictions. One of the classic aspects of non-stationarity, is the seasonality of epidemiological dynamics, linked to environment and climate [3–4] but the environmental variability can shape the disease propagation in unforeseeable ways on small and large spatial scales [5–8]. Intervention and control may also modify the course of an epidemic. A less well-described but equally important cause of non-stationarity is linked to social cycles, *e.g.* school terms, religious holidays and agricultural cycles [9–13]. Research increasingly focuses on the effect of behavioral change in the presence of epidemiological risk as a source of non-stationarity [14–16]. Societal responses and changing human behavior play an important role in our connected society. Thus, during an epidemic, depending on the availability of information on the disease, people exhibit a variety of behaviors including anxiety and social distancing that might greatly influence the course of an epidemic.

For all of the above reasons, the spread of pathogens through human populations can be complex and hard to predict. In the face of this complexity, mathematical models offer valuable tools to study the dynamics of epidemic diseases, in order to synthesize information to understand observed epidemiological patterns and to test different hypothesis on the underlying key mechanisms [17]. Moreover, mathematical models play a crucial role in infectious disease prevention by assessing the impact of different control measures, *e.g.* vaccination strategies [18–19].

Nonetheless, there are very few, if indeed any, cases where modelers can access all the necessary information to reliably predict the course of an epidemic. This is particularly the case when we consider the non-stationarity features of epidemics and their transient nature poses a challenging problem for modeling. Further to this, different hypothesis must be formulated. In the case of influenza, for example, some researchers have suggested using a quantitative relationship between climatic variables and the effective transmission rate [20]. Another recent example illustrates non-stationarity in epidemiology. Between November 2010 and February 2011, despite a low level of population susceptibility, an unexpected third wave of infection by the H1N1pdm09 pandemic virus was observed in the United Kingdom. Using a

compartmental mathematical model of influenza transmission, this third wave was explained, by a substantial increase in the transmissibility of the H1N1pdm09 virus [21]. It has been proposed that this modification of the transmissibility was caused by the virus evolution with a better adaptation to the human host, or by climatic factors, namely the very cold weather experienced in the United Kingdom at that time, or by a combination of these factors [21].

To tackle the problem of non-stationarity in epidemiology, some approaches use a linear function to reconstruct the effective reproduction number (average number of secondary cases per primary case, R_{eff}). Wallinga and Teunis [22] proposed a generic method that requires only case incidence data and the distribution of the serial interval (the time between the onset of symptoms in a primary case and the onset of symptoms of secondary cases) to estimate R_{eff} over the course of an epidemic. This approach has been improved by numerous authors and applied to real time estimation of R_{eff} [23–25]. Other authors estimated using mathematical models R_{eff} for each season [26,27]. To calculate the time-varying infection rate the reconstructed time series of R_{eff} derived from the notification data can be used [28].

However, the time evolution of R_{eff} by definition depends not only on the time evolution of the epidemiological parameters but also on the number of susceptibles. More complex approaches have therefore been proposed. These approaches use semi-mechanistic models that incorporate the known compartmental structure of disease transmission but do not specify the form of the transmission rate equation that is estimated based on the data. In an early paper, the force of infection is estimated by neuronal network or kernel regression [29]. Now it is more common to use B-spline [30–33]. An alternative approach is to use diffusion models driven by fractional Brownian motion to model time-varying parameter of major epidemiological significance [34–36]. The models developed assign diffusion processes to the time-varying parameters embedded in a state-space framework. With the Kalman filter, the time-evolution of some key parameters (average transmission rate, mean incubation rate, and basic reproduction rate) were estimated during the course of the HIV/AIDS epidemics in the Paris region [34–35]. Dureau et al. [36] generalized this approach using a Bayesian framework with an adjusted adaptive particle Markov chain Monte Carlo algorithm (PMCMC), but only applied to the transmission rate, for short epidemics, with application to the 2009 pandemic flu. Very recently, an algorithm relying on robustly estimating the time-varying infection rate, based on the method of the unknown input observers from control theory, has been proposed [37]. Similarly, an approach for the reconstruction of time-dependent transmission rates, by projecting onto a finite subspace, spanned by Legendre polynomials, has been introduced [38].

In our previous works [34–36], we have introduced an approach for reconstructing the time evolution of some key parameters with just the weak hypothesis according to which they follow a basic stochastic process. The parameter time evolution is estimated solely based on observations of the incidence or the prevalence. Here, we propose to expand this approach to recurrent epidemics over time periods longer than just one season. The underlying idea of this approach is to capture unknown influences by considering time-varying parameters. As with other semi-mechanistic approaches, the key advantage of this approach, for the parameter dynamics, is that it is data-driven, and thus the shape of change does not need to be specified beforehand. We applied our framework both to a toy model, where parameters and observation process are known, and to two real data sets. This allows us to demonstrate that this data-driven approach is very effective for tackling the non-stationarity of recurrent epidemics, even with long time series. It has other benefits too. For instance, with limited access to information, it can capture unknown influences. By so doing, and by analyzing the parameter time evolution, this framework allows a more thorough analysis of the different influences, facilitating their introduction in more complex models with pertinent hypotheses based on observations.

Models with time-varying parameters

Our approach is based on three main components: an epidemiological model embedded in a state-space framework, a diffusion process for each time-varying parameter and an up-to-date Bayesian inference technique based on adaptive PMCMC.

The main advantage of the state-space framework is the use of two sets of equations, the first set describes the propagation of the disease in the population and the second is for the observation process. This allows for consideration of unknowns and uncertainty both in the epidemiological mechanisms and in the partial observation of the disease:

$$\begin{cases} \dot{x}(t) = g(t, x(t), \theta'(t), u(t)) \\ y(t)|x(t) \sim f(h(x(t)), \theta'(t)) \end{cases} \quad (1)$$

The first equation is for the epidemiological model, with $x(t)$ representing the state variables (for instance, $S(t)$ the susceptibles, $I(t)$ the infectious and $R(t)$ the removed for the classical SIR model) and $\theta'(t)$ the epidemiological parameters. The second is the observational process defined by probabilistic law f and a reporting rate on transformation of some state variable $h(x(t))$ because we may not be able to directly measure all state variables but just some or a function of them. In these equations, $y(t)$ are partial observations of $x(t)$, $u(t)$ is the process noise describing different form of stochasticity and the observational noise is included in f . In our applications, $h(x(t))$ will be the cumulative sum of new cases over the observation time step, that is generally the quantity observed by Public Health systems.

Considering the time-varying parameters $\theta(t)$ as a subset of $\theta'(t)$, we make the assumption that they evolve more or less randomly and do not follow a defined mathematical function. In the absence of prior information the use of diffusion motion allows us to impose few restrictions on the evolution of $\theta(t)$. We consider that they follow a continuous diffusion process (a discrete diffusion process was used in [35]):

$$\begin{aligned} d\theta(t) &= \sigma dB(t) \\ \text{or} \\ d\log(\theta(t)) &= \sigma dB(t) \end{aligned} \quad (2)$$

where σ is the volatility of the Brownian process ($dB(t)$) and will be estimated during the fitting process. The use of a Brownian process can be viewed as a weak hypothesis for the imposed motion of $\theta(t)$ and the volatility σ being a regularized factor. Intuitively, the higher the values of σ the larger the changes in $\theta(t)$. The logarithm transformation avoids negative values which have no biological meaning. When prior knowledge on $\theta(t)$ is available this Brownian process can be modified to account for a drift in (2) (see [36]).

For the time-varying parameter, we focus on the parameter of the force of infection classically defined as:

$$\lambda(t) = \beta(t) \cdot \frac{S(t) \cdot I(t)}{N} \quad (3)$$

with $\beta(t)$ the transmission rate usually defined by a sinusoidal function. The control or the behavior modification can also be taken into account:

$$\lambda(t) = \beta(t) \cdot \frac{(S(t)^{\varepsilon_S(t)}) \cdot (I(t)^{\varepsilon_I(t)})}{N} \quad (4)$$

$\varepsilon^i(t)$ describe the clustering of the population [39,40] but can also describe a reduction in the population due to voluntary avoidance behavior or social distancing. However due to the

absence of structural identifiability properties [41, 42] it should be very difficult to estimate simultaneously both $\beta(t)$ and $\varepsilon^i(t)$.

For model estimation we use Bayesian methods, coupling particle filter and MCMC for partially observed stochastic non-linear systems [36,43] (see Methods). The implementation provided in SSM software [36] is used.

Results

Reconstructing dynamics and time-varying parameter for a toy SIRS model

We start our demonstration by showing that it is possible to reconstruct both the trajectory of a SIRS model (SIRS stands for Susceptibles, Infectious, Removed and Susceptibles again) and that of the sinusoidal transmission rate. In this example, the trajectory of each variable has been simulated with a model for which all the parameters were known. Moreover we also knew the observation process that has generated the data, a Poisson law for the incidence with an observation rate equal to 1. Fig 1 displays the reconstructed trajectories of both the

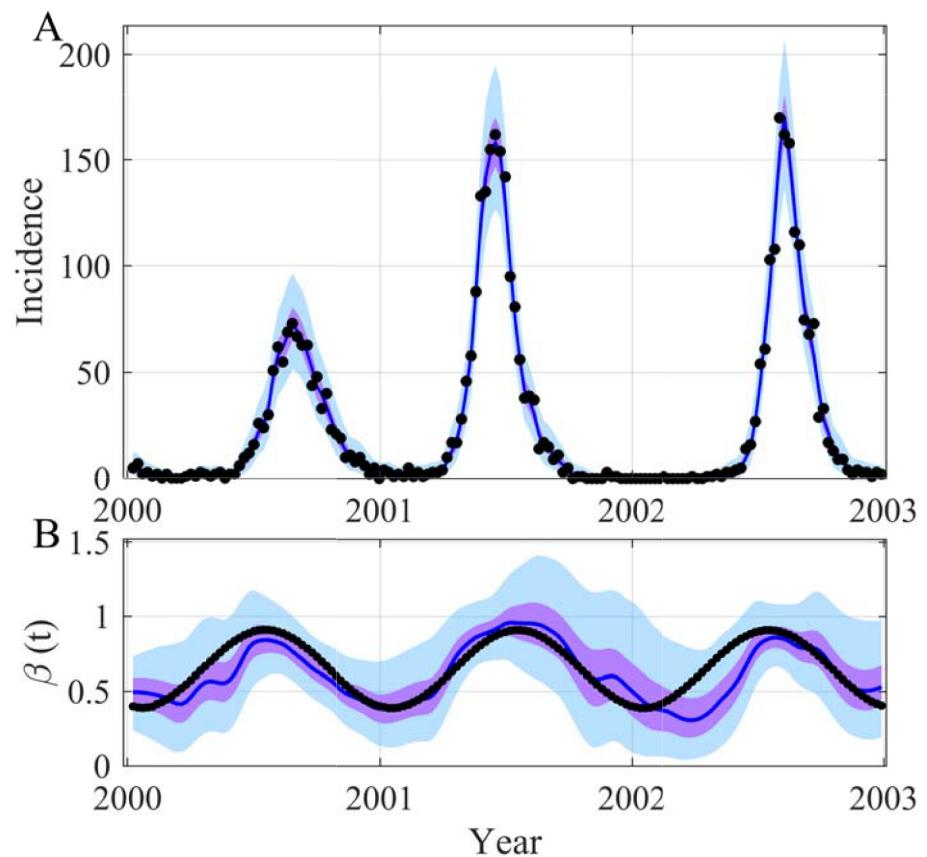


Fig 1. Reconstruction of both the incidence (A) and the time evolution $\beta(t)$ (B) for the SIRS model. In (A) the black points are observations generated with a Poisson process with a mean equal to the incidence simulated by the model. In (B) the black points are the true values of $\beta(t) = \beta_0(1 + \beta_1 \sin(2\pi t/365 + 2\pi\phi))$. The blue lines are the median of the posterior, the mauve areas are the 50% Credible Intervals (CI) and the light blue areas the 95% CI. For all the figures, the observation process is also applied to the inferred incidence trajectory. The time unit of the model is day, the initial date is arbitrary (2000-01-09) and parameters used for the SIRS model are as follows: $\mu = 1/(50*365)$, $\alpha = 1/(7*365)$, $\gamma = 1/14$, $\beta_0 = 0.65$, $\beta_1 = 0.4$, $\phi = -0.2$, $\rho = 1$, $N = 10000$, $S(0) = 600$, $I(0) = 30$. The prior and posterior distributions of the inferred parameters are in S1 Fig.

<https://doi.org/10.1371/journal.pcbi.1006211.g001>

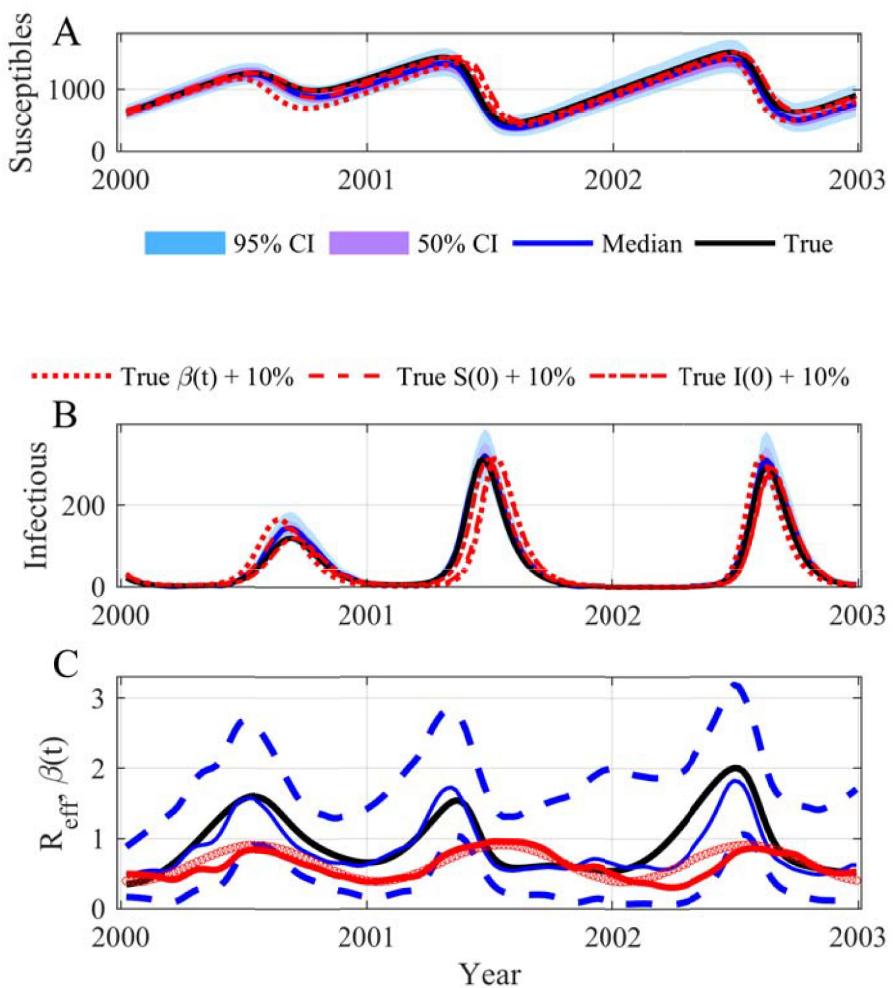


Fig 2. Simulation of the SIRS model: (A) Susceptibles; (B) Infectious; (C) Time evolution of both R_{eff} and $\beta(t)$. In (A) and (B) the black lines are the true values, the blue lines are the median of the posterior, the mauve areas are the 50% CI and the light blue areas the 95% CI. In (A) and (B) the susceptibles and infectious trajectories with a modification of 10% of the value of $S(0)$, $I(0)$ and $\beta(t)$ have been added to show the weak sensitivity of the SIRS model to these values. In (C) the black line is the true values of R_{eff} , the red dot line is the true time evolution of $\beta(t)$ and the red line the median of its posterior. Model parameters as in Fig 1: The time unit of the model is day, the initial date is arbitrary (2000-01-09) and parameters used for the SIRS model are the following: $\mu = 1/(50*365)$, $\alpha = 1/(7*365)$, $\gamma = 1/14$, $\beta_0 = 0.65$, $\beta_1 = 0.4$, $\phi = -0.2$, $\rho = 1$, $N = 10000$, $S(0) = 600$, $I(0) = 30$.

<https://doi.org/10.1371/journal.pcbi.1006211.g002>

incidence and the transmission rate highlighting the potential of the method. The parameter estimations are in perfect agreement with the values used to generate the observations and the estimation process has correctly converged ([S1](#) and [S2](#) Figs). This clearly demonstrates the feasibility of accurately ascertaining the time evolution of the transmission rate and correctly estimating the R_{eff} (see Fig 2). It is worth emphasizing that the SIRS model is a complicated example for different reasons. First, even with a constant transmission rate the SIRS model can generate oscillations (damped oscillations, see [[17,44](#)]). Secondly, the model trajectories are not very sensitive, a modification of $\pm 10\%$ can induce minor modifications of the trajectories

that are inside or near the 95% CI of our inferences ([Fig 2](#)). Moreover in this example we have used initial conditions outside the attractor of the dynamics to generate transients that appear more realistic for real applications, but are more complex to reconstruct. The robustness of our approach has also been tested: (i) using long time series and initial conditions near the attractor ([Fig 3A](#) and [S3A Fig](#)); (ii) modifying the number of inferring parameters ([S4–S6 Figs](#)), for instance estimating just the volatility parameter ([S7](#) and [S8 Figs](#)); (iii) considering the possibility of not using the transformation log in the diffusion process ([S9](#) and [S10 Figs](#)) and (iv) using a true $\beta(t)$ with 2 or 3 periodic components ([Fig 3B](#) and [3C](#) and [S3B](#) and [S3C Fig](#)).

We have also explored the performance of our approach by comparing their inferences to those of the true model. The re-estimation of the true model on its own data is displayed in [S11–S13 Figs](#). [Table 1](#) presents indices of the goodness-of-fit of the true model and models with time-varying $\beta(t)$ with different number of parameters inferred. As expected, the error on $\beta(t)$ is smaller when the true equation is used ([Table 1](#)). However, regarding the estimated incidence, the true model and our approach give similar results both in terms of mean and variance ([Table 1](#)). It could be argued that the price of the flexibility of our approach is a greater variability in some of the trajectory estimations ([Table 1](#)). Nevertheless the average dynamics are always estimated correctly.

As misspecification is an important problem (e.g. [45]) we have also compared the performance of our approach to those of a misspecified seasonal SIRS model. We have thus used the example of a sinusoidal $\beta(t)$ with two periodic components (see [Fig 3B](#)) and computed the indices of the goodness-of-fit of the true model with the SIRS model with 1 year sinusoidal $\beta(t)$ and with our time-varying periodic $\beta(t)$. The results clearly show that our approach performed better than the misspecified model for the three trajectories analyzed, Incidence, β and R_{eff} ([Table 2](#)). Once again the price of the flexibility of our approach is a greater variability in some of the trajectory estimations. However this is preferable to a large error in the median trajectories as occurred in those observed with the misspecified model ([Table 2](#)).

Our methodology is also applicable to other more complex or simpler tasks. For instance, it can follow the time evolution of a parameter describing the availability of susceptibles, $\varepsilon_S(t)$ ([Fig 4](#) and [S14](#) and [S15 Figs](#)). [Fig 4](#) shows the accurate reconstruction of the trajectory of the incidence and also of the trajectory of $\varepsilon_S(t)$ that shifted at a given time point and decreased slightly thereafter. This highlights once again the potential of our approach as it is never easy to estimate a discontinuous dynamic with a continuous process (2).

Application to real datasets: Flu

In previous works, the dynamics of influenza in Israel have been analyzed using a discrete deterministic SIRS model and weekly data from Israel's Maccabi health maintenance organization [20,46]. To describe the seasonality of this recurrent epidemic, the authors used a linear model between the transmission rate and local climatic variables, daily temperature and relative humidity [20,46]. We have re-analyzed their dataset (but limited to 1998–2003 due to a modification in the reporting) to reconstruct the time evolution of $\beta(t)$. Our results ([Fig 5](#) and [S16 Fig](#)) clearly show the potential of our method, highlighting that the $\beta(t)$ fluctuations are more irregular and complex than a simple sinusoidal function.

Application to real datasets: Dengue

Our last example is on dengue in Cambodia. Again the idea is to relax the assumption of a sinusoidal $\beta(t)$ in a SEIR model. Monthly data from the capital Phnom Penh [47], for which the meteorological data is available from the international airport, was used. We can accurately describe the 12 year time series and reconstruct the time evolution of $\beta(t)$ ([Fig 6](#) and [S17 Fig](#)).

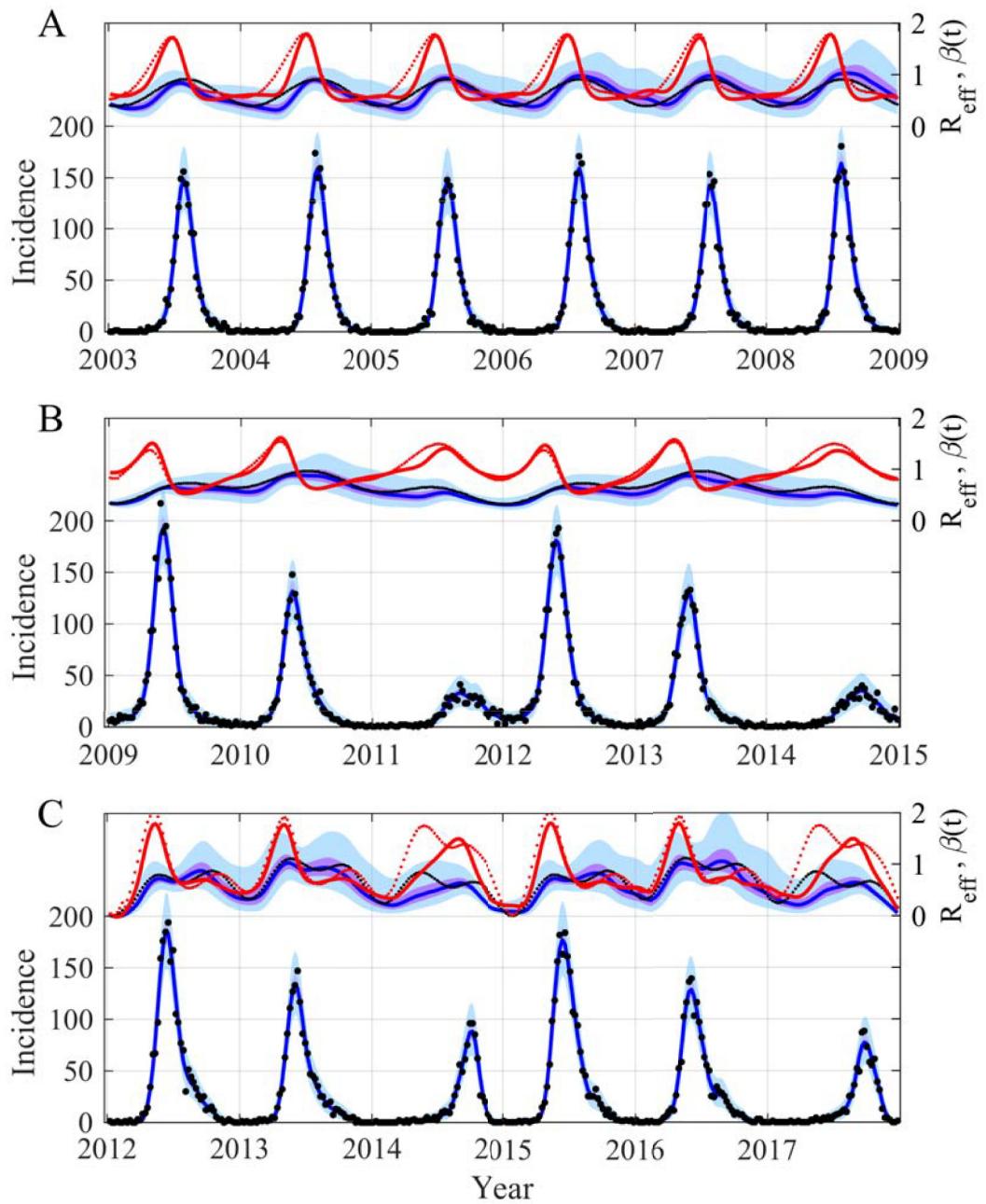


Fig 3. Reconstruction of both the incidence (bottom panel) and the time evolution of $\beta(t)$ and R_{eff} (top panel) for a SIRS model with the initial conditions near the attractor of the dynamics. The true β is generated by $\beta(t) = \beta_0(1 + \beta_1 \sin(2\pi t/365 + 2\pi\phi) + \beta_2 \sin(2\pi t/(3365) + 2\pi\phi)) + \beta_3 \sin(2\pi t/(0.5365) + 2\pi\phi))$ with in (A) $\beta_1 = 0.4, \beta_2 = 0, \beta_3 = 0, S(0) = 911.5, I(0) = 3.5$; in (B) $\beta_1 = 0.4, \beta_2 = 0.3, \beta_3 = 0, S(0) = 1735, I(0) = 20$; and in (C) $\beta_1 = 0.1, \beta_2 = 0.1, \beta_3 = 0.1, S(0) = 3365, I(0) = 3$. The other parameters used are as follows: $\mu = 1/(50*365), \alpha = 1/(7*365), \gamma = 1/14, \beta_0 = 0.65, \phi = -0.2, \rho = 1, N = 10000$. In both panels blue lines are the median of the posterior, the mauve areas are the 50% CI and the light blue areas the 95% CI. In the top panel, the red line is the reconstructed R_{eff} , the points are the true values of R_{eff} (red) and of $\beta(t)$ (black). In the bottom panel the black points are observations generated with a Poisson process. The prior and posterior distributions of the inferred parameters are in S3 Fig.

<https://doi.org/10.1371/journal.pcbi.1006211.g003>

Table 1. Comparison of goodness-of fit indices for different models and different numbers of parameters inferred. The indices are computed on Incidence, β and R_{eff} trajectories: RMSE: root mean square error using the median; MAPE: maximum absolute percentage error using the median; MIQR: mean inter-quartile range. The parameter values used are in the captions of the figures. For comparison purposes we used a stochastic version of the SIRS model with sinusoidal β and for all figures the observation process is applied to the inferred incidence trajectory.

Model		Sinusoidal β	Time-varying $\beta(t)$	Time-varying $\beta(t)$	Time-varying $\beta(t)$
Figure		S11 Fig	Fig 1	S4 Fig	S7 Fig
# of parameters inferred		8	7	5	1
Incidence	RMSE	4.18	4.14	4.13	4.12
	MAPE	2.24	2.34	2.39	2.24
	MIQR	1.03	0.95	1.04	1.03
$\beta(t)$	RMSE	0.07	0.11	0.14	0.13
	MAPE	0.29	0.48	0.52	0.50
	MIQR	0.07	0.43	0.36	0.33
R_{eff}	RMSE	0.06	0.19	0.24	0.24
	MAPE	0.18	0.47	0.50	0.50
	MIQR	0.23	0.64	0.53	0.41

<https://doi.org/10.1371/journal.pcbi.1006211.t001>

Our results stress that the $\beta(t)$ oscillations are more complex than a simple sinusoidal function. Sometimes bi-modality occurs over one season. In general one observes a fast growth of $\beta(t)$ and a slow decrease. Moreover the amplitude of the $\beta(t)$ varies from year to year, perhaps depending on the fluctuations in the mosquito population and in the environment. Interestingly the peak in $\beta(t)$ appears 1 to 2 months before the incidence peak. This delay can be explained by the extrinsic incubation period and might be used in a warning system.

To explain the $\beta(t)$ oscillations we have explored the potential effects of local and global climatic variables using wavelet decomposition [48] as one of our main underlying hypotheses is non-stationarity. We observed very significant coherency between $\beta(t)$ and climate for the local climate for the seasonal mode (Fig 7 and S18–S20 Figs) and also for the 2–3 year components with global climatic variable (S21 Fig). Thus, the rhythm of $\beta(t)$ can be explained perfectly by climatic factors. Nevertheless, again mainly due to large non-stationarity, by using solely one or two climatic variables we are able to correctly describe dengue evolution in the short-term (Fig 7C, red area) but not over a large time period (Fig 7C, blue area). This reflects

Table 2. Comparison of goodness-of fit indices between a misspecified model and a model with a time-varying parameter. The reference is a sinusoidal model with 2 periodic components. The parameter values used are in the captions of Fig 3. The indices are computed on Incidence, β and R_{eff} trajectories: RMSE: root mean square error using the median; MAPE: maximum absolute percentage error using the median; MIQR: mean inter-quartile range. For comparison purposes we used a stochastic version of the SIRS models with sinusoidal β and for all figures the observation process is applied to the inferred incidence trajectory.

Model		Reference: Sinusoidal β with 2 periods (see Fig 3B caption)	Sinusoidal β with 1 year period (see Fig 3A caption)	Time-varying $\beta(t)$ (5)
# of parameters inferred		8 including 4 for β	7 including 3 for β	5 including 1 (σ) for β
Incidence	RMSE	4.80	5.21	4.83
	MAPE	2.21	1.98	2.40
	MIQR	0.87	1.03	0.98
$\beta(t)$	RMSE	0.05	0.26	0.08
	MAPE	0.22	1.43	0.27
	MIQR	0.06	0.06	0.28
R_{eff}	RMSE	0.06	0.22	0.08
	MAPE	0.12	0.44	0.20
	MIQR	0.20	0.20	0.46

<https://doi.org/10.1371/journal.pcbi.1006211.t002>

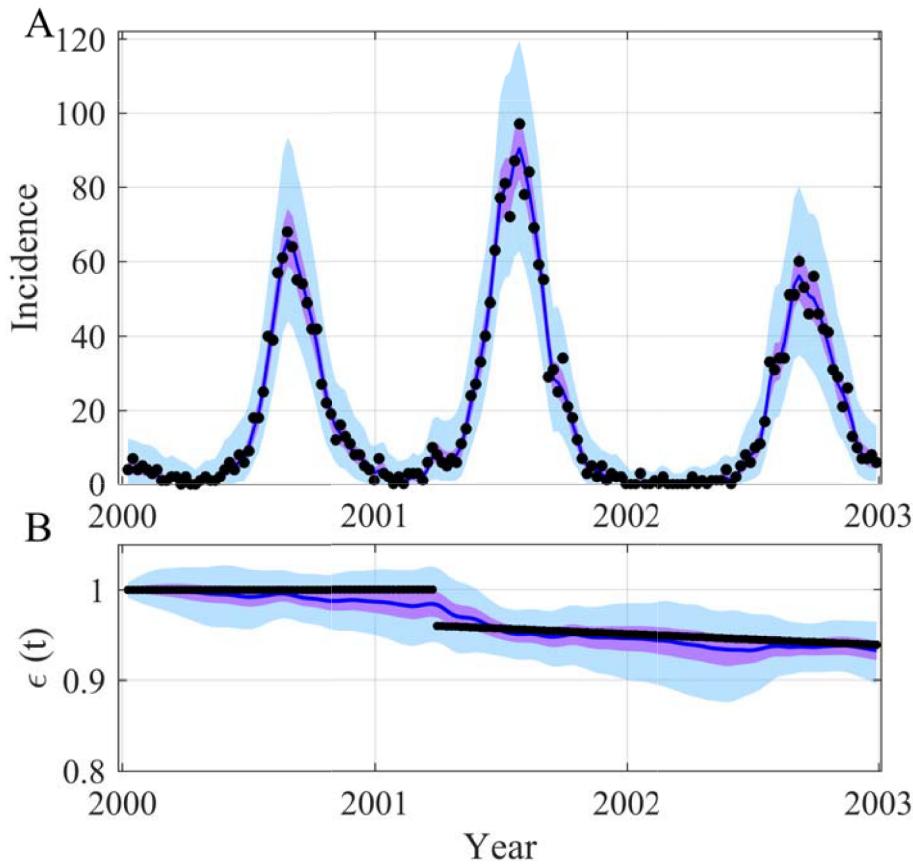


Fig 4. Reconstruction of both the incidence (A) and the time evolution of $\varepsilon_s(t)$ (B) for the SIRS model. In (A) the black points are observations generated with a Poisson process with a mean equal to the incidence simulated by the model. In (B) the black points are the true value of $\varepsilon_s(t)$: $\varepsilon_s(t) = 0.96 - (0.012/365)*(t-t_{shift})$ at $t_{shift} = 450$ days after t_0 in days. The blue lines are the median of the posterior, the mauve areas are the 50% CI and the light blue areas the 95% CI. The time unit of the model is day, the initial date is arbitrary (2000-01-09), $\beta(t) = \beta_0(1 + \beta_1 \sin(2\pi t/365 + 2\pi\phi))$ and parameters used for the SIRS model are as follows: $\mu = 1/(50*365)$, $\alpha = 1/(7*365)$, $\gamma = 1/14$, $\beta_0 = 0.65$, $\beta_1 = 0.4$, $\phi = -0.2$, $\rho = 1$, $N = 10000$, $S(0) = 600$, $I(0) = 30$. The prior and posterior distributions of the inferred parameters are in S15 Fig.

<https://doi.org/10.1371/journal.pcbi.1006211.g004>

the complexity of such a disease where the ecology of the vectors, the environment, the climate, the immune status of the human population and its behavior are all involved. This large non-stationarity association between dengue and climatic factors has recently been demonstrated using statistical models (dynamic generalized linear models) and data from a medium-sized city in Colombia [49]. The authors showed that dengue cases correlate with climatic variables (temperature, rainfall, solar radiation and relative humidity) but these correlations change over time, some intervals showing a positive association, while in others the association is negative [49]. The non-stationarity association between dengue and climate may be explained by the fact that a climatic variable has different effects depending on the biological cycle of the pathogen or of the vector. Moreover the effects of one climatic variable can also depend on other climatic variables potentially enhancing the non-stationarity association.

Discussion

As there remain numerous uncertainties during the course of each epidemic, we are increasingly aware of the importance of developing adequate statistical and mathematical tools. Such

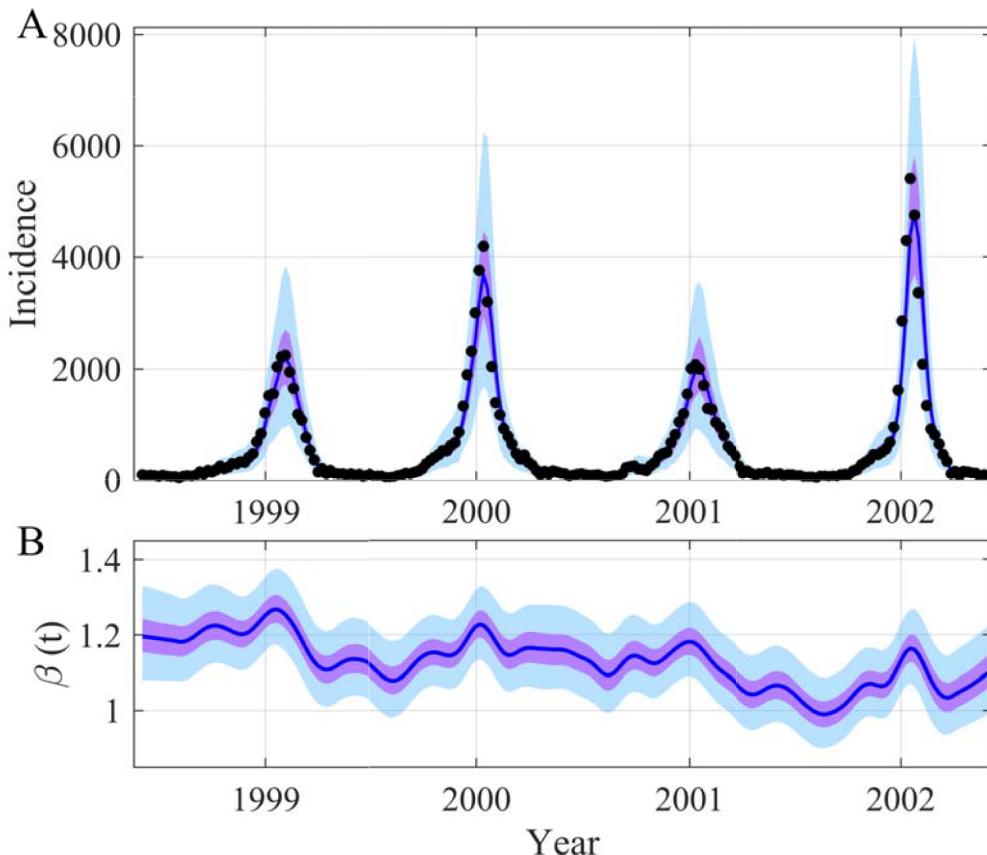


Fig 5. Reconstruction of both the incidence (A) and the time evolution $\beta(t)$ (B) in the case of the 1998–2002 seasonal flu epidemics in Israel. In (A) the black points are influenza-like illness incidence collected by Israel's Maccabi health maintenance organization [46]. The blue lines are the median of the posterior, the mauve areas are the 50% CI and the light blue areas the 95% CI. The prior and posterior distributions of the inferred parameters are in S16 Fig.

<https://doi.org/10.1371/journal.pcbi.1006211.g005>

tools need to take account of the time-varying nature of the underlying ecological and biological mechanisms as well as social and behavioral influences involved in an epidemic. Because of this, time-varying parameters modeled with a diffusion process, that track epidemiological patterns and update the key parameters according to data appear to be a worthwhile approach. Indeed developing a more complex model would be difficult considering the relative paucity of available data.

We propose a flexible modeling framework that encompasses time-varying aspects of the epidemic. It does this via diffusion process equations for time-varying parameters and also considers uncertainty associated with key parameters and data. This data-driven framework for time-varying parameters has been coupled with simple stochastic models and a robust Bayesian procedure for inference. To test its efficiency, our proposed methodology was first applied to a toy model and then to real epidemiological examples.

Our results clearly demonstrate the potential of our framework. Firstly, our methodology was able to accurately reconstruct both the incidence and the sinusoidal transmission rate of a simple SIRS model just based on noisy observations (Figs 1–4 and S4,S5,S7,S9 and S14 Figs). Based on these reconstructions one can also closely estimate R_{eff} which is one of the key relevant epidemiological parameters. Our results also highlight the flexibility of our developed methodology. It can reconstruct the time evolution of a shifting parameter ($\varepsilon_S(t)$, see Fig 4 and

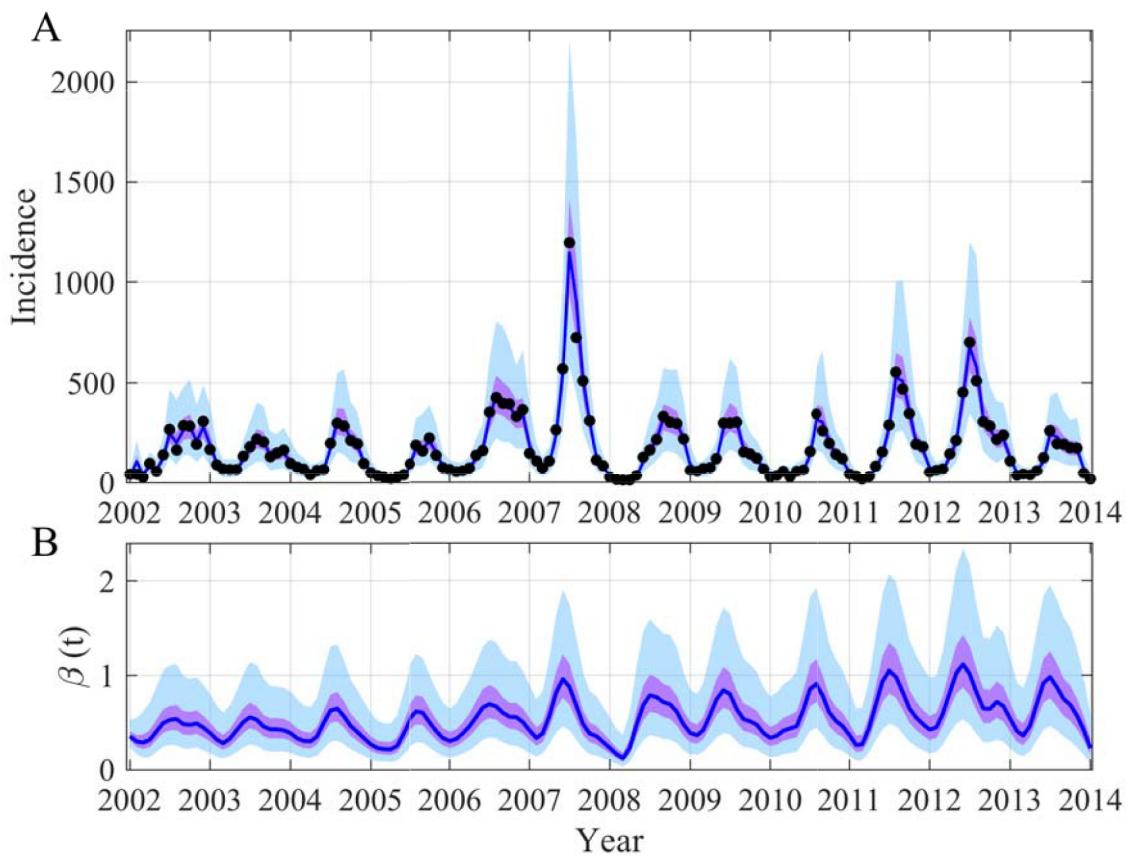


Fig 6. Reconstruction of both the incidence (A) and the time evolution of $\beta(t)$ (B) in the case of the 2002–2013 dengue epidemics in the province of Phnom Penh (Cambodia). In (A) the black points are dengue incidence recorded by the Cambodian National surveillance (National Dengue Control Program from the Ministry of Health, see [47]). The blue lines are the median of the posterior, the mauve areas are the 50% CI and the light blue areas the 95% CI. The prior and posterior distributions of the inferred parameters are in S17 Fig.

<https://doi.org/10.1371/journal.pcbi.1006211.g006>

S14 Fig) as well as an oscillating parameter that influences the nonlinear part of the model ($\beta(t)$, see Figs 1–3 and S4,S5,S7 and S9 Figs). The comparison using goodness-of-fit indices with the inferred true model allows us to highlight the fact that our methodology performs as well for the observed incidence. Its flexibility results in greater variability in some other trajectories mainly $\beta(t)$ and R_{eff} (Table 1). Moreover, in the absence of knowledge of the true evolution of the transmission rate, our approach appears to capture the dynamic observed more accurately than a misspecified model (Table 2). Secondly, applied to real datasets, our framework is able, based solely on simple stochastic models, to reconstruct complex epidemics such as flu or dengue over long time periods (Figs 5 and 6). In such cases, the reconstruction of the time evolution of the transmission rate clearly stresses that, on real datasets, it is difficult to assimilate the dynamic of this parameter as a simple sinusoidal function. It is more irregular in amplitude and sometimes multi-modal over one season.

Considering the paucity of information available regarding the complexity of the mechanisms involved during an epidemic, describing and fitting a full model for a given transmissible disease is always challenging. Our data-driven methodology can be used as a first step towards a better understanding of a complex epidemic, where data is limited or lacks certainty. Indeed most of the unknowns and uncertainties can be put into time-varying parameters. The

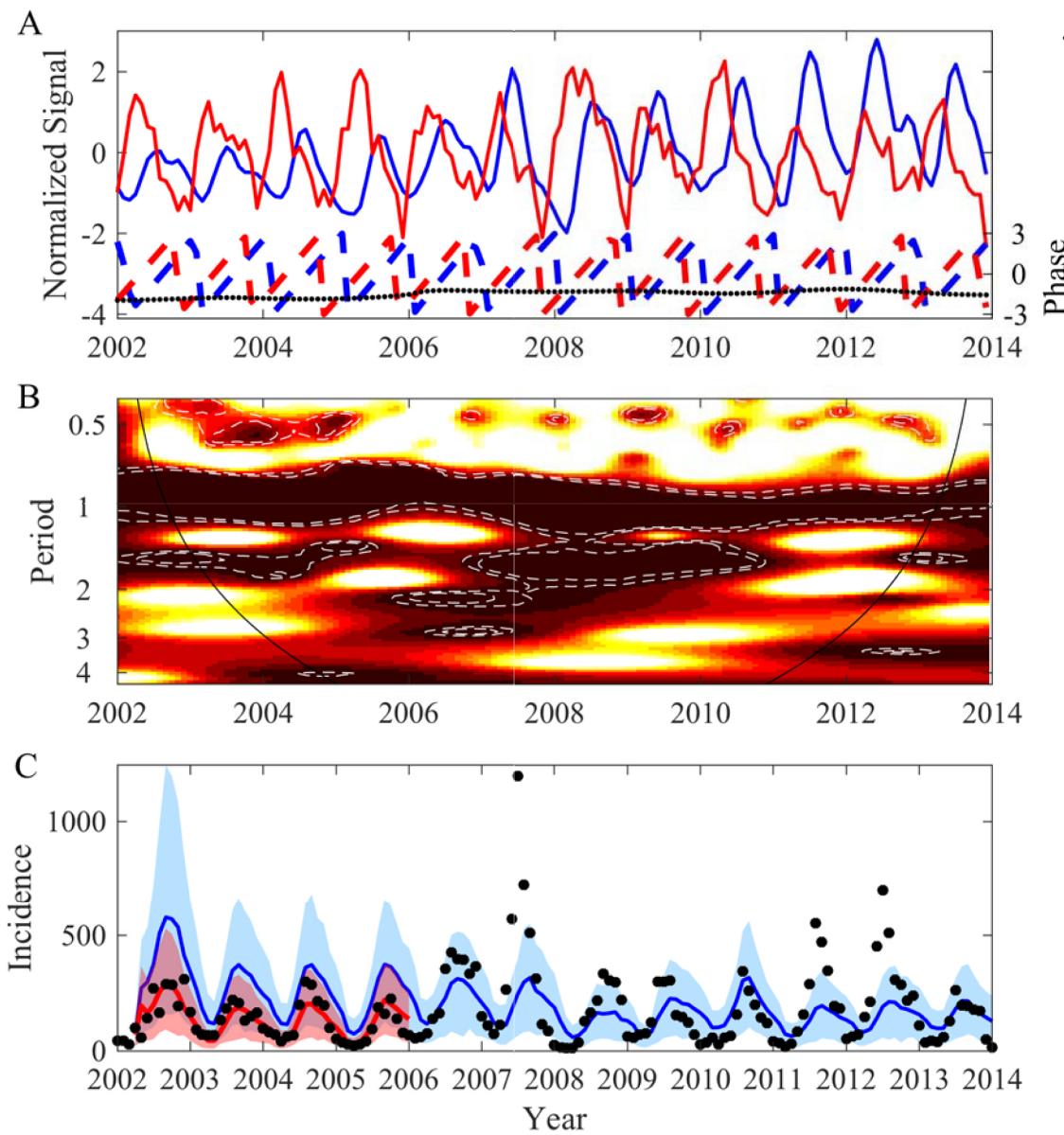


Fig 7. Association between dengue transmission rate and monthly average maximum temperature in the province of Phnom Penh (Cambodia). (A) Time evolution of the normalized median of $\beta(t)$ (blue line) and average temperature (red line) as well as the evolution of their phase computed based on wavelet decomposition (see **Method** and [48,70]), blue dashed line for the normalized $\beta(t)$, red dashed line for the normalized averaged temperature and black dotted line for their phase difference. (B) Wavelet coherence (see **Method** and [48,70]) between the reconstructed $\beta(t)$ and average temperature. The colors code for low values in white to high values in dark red. The white dashed lines show the 90% and the 95% CI computed with adapted bootstrapping [71]. (C) Model simulations using a linear model describing $\beta(t)$ with monthly average maximum temperature (S18 Fig) and monthly average minimum temperature (S19 Fig) ($\beta(t) = a_0 + a_1 \text{MaxTemp}(t) + a_2 \text{MinTemp}(t)$). The red line is the median of the posterior and the red area is the 95% CI when parameters are estimated for the period 2002–2005. The blue line is the median of the posterior and the light blue area is the 95% CI when parameters are estimated for the full time period. The black points are dengue incidence recorded by the Cambodian National surveillance [47].

<https://doi.org/10.1371/journal.pcbi.1006211.g007>

potential effects of all these uncertainties can then be explored by analyzing the reconstructed time evolution of the time-varying parameters. See Fig 7 for such preliminary analysis of dengue in Phnom Penh. This allows a more thorough analysis of the influences and the

interactions between both the human behavior and complex environmental drivers. In a recent paper [50], the authors reviewed evidence of interactions between seasonal influenza virus and other pathogens (bacteria or virus). They concluded that it is important to incorporate these different coinfecting pathogens in models of seasonal flu in order to get a better estimate of the burden of influenza. Our framework could be an alternative to the development of complex models with all the potential interactions between pathogens and to estimate the strength of the interactions. After reconstructing the time evolution of the transmission rate the statistical association between the coinfecting pathogens and the transmission rate could be tested. This screening may facilitate the construction of more complex models that could incorporate only the most significant coinfecting pathogens in the seasonal flu model.

Our methodology also has other advantages. Taking account of the simplicity of the model used, and the fact that weak hypotheses on the dynamics of the time-varying parameters have been included, our proposed methodology can retrospectively test the impact of interventions. This has previously been done in the case of HIV epidemics [34–35], where it was hypothesized that the reduction in the transmissibility was due to a modification of the sexual behavior in the population and the increase in the seropositive period duration due to the introduction of the first antiviral treatments. Evaluation of interventions has also been done recently in the case of the Ebola epidemic in West Africa [51]. The relative simplicity of our methodology is also suitable for short-term predictions and it can then easily be used to predict an epidemic in real time. Starting with a given estimated state defining the system, the fitting process can be run again each time new data is available and the new posteriors are used for new predictions [36]. This can inform public health decisions and indeed has been done recently to great effect in the case of the Ebola epidemics in West Africa [52].

A major challenge in model fitting is the reliability of data collected and also the non-identifiability of the mechanistic models that always have very rich dynamical behavior. The question of identifiability is too often avoided in epidemiological models applied to a topical Public Health issues. There is, however, considerable literature on this subject (e.g. [41,42,53–55]). Identifiability is not evident even for a simple seasonal SIR model [56]. To solve this problem one can fit a combination of parameters or fix some of them (the population size for instance) [57]. In our applications there is a clear limitation due to practical non-identifiability of reporting rate and initial conditions. To fix these problems we have used informative priors (see [Method](#)). Using informative priors or fixing some parameters gives very similar results (compare [Fig 1](#) and [S4–S7 Figs](#)). Related to this is the misspecification of models [45]. In our cases, as with other semi-mechanistic models the time-varying parameter methodology captures some of the information in the data but not in the mechanistic part of the model. If the model is misspecified due to lack of precision, it compensates for it and the dynamics of $\beta(t)$ will drive improvements in the model to make it more complex and realistic ([Table 2](#)). If the model is misspecified to the extent that it creates mechanisms that do not exist, the reconstructed $\beta(t)$ would compensate for these effects but it will be harder to interpret.

In this work we have used simple mechanistic models. The proposed methodology is not limited to simple models. For instance, a two-strain dengue model has also been tested. In this case the main problem was linked to the unavailability of both seroprevalence and incidence for each strain. Indeed, one of the major difficulties with these multi-strain models is the identification of the initial conditions (e.g. [58]). Nevertheless it is worth emphasizing that the Bayesian inference method used in our framework, PMCMC, the approximation of the likelihood is limited for a large number of parameters and/or equations [59]. In such cases testing other methodologies like ABC [60,61] is advisable.

It is always difficult to fit complex models with rich behaviors based on very limited information. In this regard we agree with Metcalf et al. [62] who stressed that nowadays we need

seroprevalence studies to quantify the immunological status of the population, because in most cases the magnitude of the outbreak is difficult to evaluate without precise seroprevalence data.

To tackle the uncertainty and the non-stationarity of epidemics, our methodology, although it appears non-standard, makes important progress towards a better understanding of the mechanisms responsible for disease propagation. We believe that, should it form part of the development of the next predictive tools for Public Health, it will make a significant contribution to improving the understanding and control of infectious diseases in our increasingly uncertain world.

Methods

Models

SIRS model. Our first model is a classical SIRS model with an observation rate $\rho = 1$ and Poisson law as the observational process:

$$\begin{aligned}\dot{S} &= \mu.(N - S) - \beta(t).S.I/N + \alpha.R \\ \dot{I} &= \beta(t).S.I/N - (\gamma + \mu).I \\ \dot{C} &= \beta(t).S.I/N \\ \dot{R} &= \gamma.I - (\alpha + \mu).R\end{aligned}\tag{5}$$

where S , I and R are the susceptibles, the infectious and the removed respectively, the transmission rate $\beta(t) = \beta_0(1 + \beta_1 \sin(2\pi t/365 + 2\pi\phi))$, $1/\alpha$ is the average duration of immunity, γ is the recovery rate and μ is the recruitment or mortality rate. In (5), C is the number of new cases, then $h(x(t))$ is the cumulative sum of $C(t)$ over the observation time step, 7 days. With this model $R_{eff}(t) = \beta(t).S(t)/(N.\gamma)$. The parameter values are in the caption of Fig 1. For the fit of our simulated data, Gaussian priors are used for epidemiological parameters (α and γ). Initially non-informative priors were used for the volatility σ , the reporting rate ρ and the initial conditions $\gamma(0)$ by $\beta(0)$ but to reduce problems linked to practical non-identifiability materialized by correlation between some estimates, informative priors were used for ρ (see S1 Fig). Some other simulations have been done fixing $\beta(0)$ and ρ or just inferring σ (see S4–S8 Figs).

SIRS flu model. For analyzing Israel flu data we have used a continuous SIRS model identical to (5), we simply added imported infectious people i in the force of infection:

$$\lambda(t) = \beta(t) \cdot \frac{S(t).(I(t) + i)}{N}\tag{6}$$

The initial guess values for the parameters are from [46]. In this example the observational process is a Negative-Binomial law with an over-dispersion parameter equal to 0.05 and the reporting rate $\rho = 0.15$ [46]. For the fit, Gaussian priors are used for epidemiological parameters (i , α , γ) and non-informative priors for the volatility σ and the initial conditions ($S(0)$, $I(0)$). S16 Fig displays the priors and the posteriors.

SEIR one strain dengue model. To describe the dengue epidemics, taking account of the available data for Phnom Penh, we have fitted a one strain model using a SEIR model:

$$\begin{aligned}\dot{S} &= \mu.(N - S) - \beta(t).S.(I + i)/N \\ \dot{E} &= \beta(t).S.(I + i)/N - (\delta + \mu).E \\ \dot{I} &= \delta.E - (\gamma + \mu).I \\ \dot{R} &= \gamma.I - \mu.R\end{aligned}\tag{7}$$

where E is for infected but not yet infectious, $\beta(t)$ is the transmission rate, $1/\delta$ is the average duration of the latent period, γ is the recovery rate and μ is the recruitment or mortality rate. The initial guess for parameter values comes from the literature [63]. In this example the observational process is a Negative-Binomial law with an over-dispersion parameter equal to 0.05 and the reporting rate ρ has been estimated using a narrow Gaussian prior. Non-informative priors are used for the volatility σ , initial condition for infected $E(0)$ and imported infectious people i . Gaussian priors are used for other parameters and initial conditions. When $E(0)$ is fitted, $I(0)$ is estimated as a steady-state value $I(0) = \delta \cdot E(0) / (\gamma + \mu)$. S17 Fig. displays the priors and the posteriors.

Inference

Stochastic framework. Due to the use of a diffusion Eq (2) for the dynamic of the time-varying parameters, the stochastic versions of the previous models have been fitted. Thus the models are considered in a stochastic framework in which the compartments are discrete and the number of reactions occurring in a time interval dt is approximated by a multinomial distribution. It is fully described in [64,65].

SMC algorithm as implemented in SSM [65]. In a model with n observations and J particles. L is the model likelihood $p(y_{1:n}|\theta)$. $W_k^{(j)}$ is the weight and $x_k^{(j)}$ is the state associated to particle j at iteration k .

1. Set $L = 1$, $W_0^{(j)} = 1/J$.
2. Sample $(x_0^{(j)})_{j=1:J}$ from $p(x_0 | \theta)$
3. **for** ($k = 0 : n-1$) **do**
4. **for** ($j = 1 : J$) **do**
5. Sample $x_{k+1}^{(j)}$ from $p(x_{k+1} | x_k^{(j)}, \theta)$
6. Set $\alpha^{(j)} = p(y_{k+1} | x_{k+1}^{(j)}, \theta)$
7. **end for**
8. Set $W_{k+1}^{(j)} = \alpha^{(j)} / \sum_{l=1}^J \alpha^{(l)}$ and $L = L \cdot \sum_{l=1}^J \alpha^{(l)}$
9. Resample $(x_{0:k+1}^{(j)})_{j=1:J}$ from $W_{k+1}^{(j)}$.
10. **end for**

Estimation with particle Markov Chain Monte Carlo (PMCMC). Since the epidemiological propagation models are considered in a stochastic framework, their likelihood is intractable and it is estimated with particle filtering methods (Sequential Monte Carlo, SMC). With a given set of parameters, the SMC algorithm reconstructs sequentially the trajectory of the state variables and the time-varying parameters, and computes the associated likelihood.

Firstly, the distribution of the initial conditions of the system is approximated with a sample of particles. Then, at each iteration, the particles are projected according to the propagation model up to the next observation point, they receive a weight reflecting the quality of their prediction compared to the observation and the total likelihood is updated. A resampling step using the weights is performed before the next iteration, in order to discard the trajectories associated with low weight particles.

In order to estimate the parameters of the system, the particle filter is embedded in a Markov Chain Monte Carlo framework, leading to the PMCMC algorithm [43]. More precisely, the likelihood estimated by SMC is used in a Metropolis Hasting scheme (particle marginal Metropolis Hastings) [43]. The proposal distribution is a Gaussian whose co-variance matrix is adapted following the framework described in [65].

The starting point of the MCMC chain is initialized using optimal values obtained from the KSimplex algorithm on a large number of parameter sets. Then, a pre-adaptation of the proposal co-variance matrix is performed with Kalman MCMC (KMCMC). Each time the idea

relies on less computationally intense algorithms in order to facilitate the exploration of parameter space. But as we use stochastic models we approximate the likelihood using the extended Kalman filter both in the simplex algorithm (KSimplex) [65] and in the MCMC (KMCMC) [36]. Then the adaptive PMCMC is executed on the output of the KMCMC with 100000 iterations and 10000 particles for the final figures. For instance, the results such as those of Fig 1 take less than 24 hours, on a blade server from PowerEdge M-Series with 40 processor cores.

PMCMC algorithm as implemented in SSM [65]. In a model with n observations and J particles.

1. $q(\cdot | \theta^{(i)})$ is the transition kernel
2. Initialize $\theta^{(0)}$
3. Using SMC algorithm, compute $\hat{p}(y_{1:n} | \theta^{(0)})$ and sample $x_{0:n}^{(0)}$ from $\hat{p}(x_{0:n} | y_{1:n}, \theta^{(0)})$
4. **for** ($i = 0 : N-1$) **do**
5. Sample θ^* from $q(\cdot | \theta^{(i)})$
6. Using SMC algorithm, compute $L(\theta^*) = \hat{p}(y_{1:n} | \theta^*)$ and sample $x_{0:n}^*$ from $\hat{p}(x_{0:n} | y_{1:n}, \theta^*)$
7. Accept θ^* (and $x_{0:n}^*$) with probability $1 \wedge \frac{L(\theta^*) p(\theta^*) q(\theta^{(i)} | \theta^*)}{L(\theta^{(i)}) p(\theta^{(i)}) q(\theta^* | \theta^{(i)})}$
8. If accepted, $\theta^{(i+1)} = \theta^*$ and $x_{0:n}^{(i+1)} = x_{0:n}^*$. Otherwise, $\theta^{(i+1)} = \theta^{(i)}$ and $x_{0:n}^{(i+1)} = x_{0:n}^{(i)}$.
9. **end for**

In order to assess convergence of the chain, the visual inspection of the chains (e.g. S2 Fig or S8 Fig) was complemented by diagnosis provided in the Coda package in R [66]. Due to the large computational cost of the algorithm, we did not run multiple independent chains, rather we relied on diagnosis using one MCMC chain and testing its stationarity: Geweke diagnosis [67] and Heidelberger and Welch's diagnosis [68]. The results are presented in S1 and S2 Tables.

Wavelet analysis

Among the various approaches developed to study nonstationary data, wavelet analysis is probably the most efficient. In particular, this method gives us the possibility of investigating and quantifying the evolution in time of the periodic components of a time series (see [69]). Wavelets constitute a family of functions derived from a single function, the “mother wavelet”, $\Psi_{a,\tau}(t)$, that can be expressed as the function of two parameters, one for the time position τ , and the other for the scale of the wavelets a , related to the frequency. More explicitly, wavelets are defined as:

$$\Psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right)$$

The wavelet transform of a time series $x(t)$ with respect to a chosen mother wavelet is performed as follows:

$$W_x(a, \tau) = \frac{1}{\sqrt{a}} \cdot \int_{-\infty}^{+\infty} x(t) \cdot \Psi^*\left(\frac{t-\tau}{a}\right) dt = \int_{-\infty}^{+\infty} x(t) \cdot \Psi_{a,\tau}^* dt$$

where $*$ denotes the complex conjugate form. The wavelet transform $W_x(a, \tau)$ represents the contribution of the scale a to the signal at different time positions τ . The computation of the wavelet transform is done along the signal $x(t)$ simply by increasing the parameter τ over a range of scales a until all coherent structures within the signal can be identified. Here, as mother wavelet, we have used the Morlet wavelet [69].

With the wavelet approach, we can estimate the repartition of variance at different scale a and different time location τ . This is known as the wavelet power spectrum: $S_x(a, \tau) = |W_x(a, \tau)|^2$.

An important point with the continuous wavelet is that the relationship between the wavelet frequency f_0 and the wavelet scale a can be derived analytically. For the Morlet wavelet this relationship is given by:

$$\frac{1}{f} = \frac{4\pi a}{f_0 + \sqrt{2 + f_0^2}}$$

Then when $f_0 = 2\pi$, the wavelet scale a is inversely related to the frequency, $f \approx 1/a$. This greatly simplifies the interpretation of the wavelet analysis and one can replace, on all equations, the scale a by the frequency f or the period $1/f$.

To determine the statistical relationship between two time series, wavelet coherence can be computed (e.g. [48,70]):

$$R_{x,y}(f, \tau) = \left(\frac{|\langle W_{x,y}(f, \tau) \rangle|^2}{|\langle W_x(f, \tau) \rangle|^2 \cdot |\langle W_y(f, \tau) \rangle|^2} \right)^{1/2}$$

where the angle brackets around terms indicate smoothing in both time and frequency, $W_x(f, \tau)$ is the wavelet transform of series $x(t)$, $W_y(f, \tau)$ is the wavelet transform of series $y(t)$, and $W_{x,y}(f, \tau)$ is the cross-wavelet spectrum. The values of wavelet coherence are between $0 < R_{x,y}(f, \tau) < 1$. The wavelet coherency is equal to 1 when there is a perfect linear relation at particular time and scale between the two signals, and equal to 0 if $x(t)$ and $y(t)$ are independent.

To complement this, phase analysis can be used to characterise the association between signals (e.g. [48,70]). The phase difference provides information on the sign of the relationship (i.e., in phase or out of phase) and can be computed, for complex mother wavelet, with the wavelet transform $W_x(f, \tau)$ as:

$$\phi_x(f, \tau) = \tan^{-1} \frac{\text{Im}(W_x(f, \tau))}{\text{Re}(W_x(f, \tau))}$$

Similarly with the cross-wavelet transform $W_{x,y}(f, \tau)$ the phase difference between the two time series can be computed:

$$\phi_{x,y}(f, \tau) = \tan^{-1} \frac{\text{Im}(W_{x,y}(f, \tau))}{\text{Re}(W_{x,y}(f, \tau))}$$

Supporting information

S1 Code. Examples of model files and code for using SSM.
(ZIP)

S1 Table. Test of the MCMC chains: Geweke diagnosis [67] that tests for the non-stationarity of the chains, the parameter means computed using the first 10% and the last 50% of the chain are compared through a Z-score (stationarity is not rejected if the Z-scores are below the critical values at 5%, NS (non-significant) in the Table). The test was implemented with the Coda package in R [66].
(PDF)

S2 Table. Test of the MCMC chains: Heidelberger and Welch's diagnosis [68] that tests for the non-stationarity of the chains (NS (non-significant) in the Table meaning stationarity is not rejected at the 5% level). The test was implemented with the Coda package in R [66].
(PDF)

S1 Fig. Prior and posterior distributions for the SIRS model inferences of Fig 1. $I(0)$, $S(0)$ initial values, $\beta(0)$ initial value of $\beta(t)$, $1/\alpha$ is the average duration of immunity, γ is the recovery rate, ρ is the reporting rate and σ is the volatility of the Brownian process of $\beta(t)$. The blue distributions are the priors and the discrete histograms are the posteriors. The medians of the prior distributions for $I(0)$, $S(0)$, $\beta(0)$, $1/\alpha$, $1/\gamma$ and ρ are the “true values” used for the simulations of the observed incidences.

(PDF)

S2 Fig. The traces of the MCMC chain for the SIRS model inferences of Fig 1. $I(0)$, $S(0)$ initial values, $\beta(0)$ initial value of $\beta(t)$, $1/\alpha$ is the average duration of immunity, γ is the recovery rate, ρ is the reporting rate and σ is the volatility of the Brownian process of $\beta(t)$.

(PDF)

S3 Fig. Prior and posterior distributions for the SIRS model inferences displayed on Fig 3 when the initial conditions are near the attractor of the dynamics. A/ Observed data generated with a SIRS model and a sinusoidal β with 1 periodic component (5). B/ Observed data generated with a SIRS model and a sinusoidal β with 2 periodic components. In A/ and B/, $I(0)$, $S(0)$ are initial values, $1/\alpha$ is the average duration of immunity, γ is the recovery rate and σ is the volatility of the Brownian process of $\beta(t)$. C/ Observed data generated with a SIRS model and a sinusoidal β with 3 periodic components, σ the volatility of the Brownian process of $\beta(t)$ is the only parameter inferred. The blue distributions are the priors and the discrete histograms are the posteriors. The medians of the prior distributions are the “true values” used for the simulations of the observed incidences.

(PDF)

S4 Fig. Reconstruction of both the incidence (A) and the time evolution $\beta(t)$ (B) for the SIRS model as in Fig 1 but only 5 parameters have been inferred, $\beta(0)$ and ρ were fixed. Model parameters as in Fig 1 and S6 Fig.

(PDF)

S5 Fig. Simulation of the SIRS model when the initial conditions are near the attractor of the dynamics and just 5 parameters inferred: (A) Susceptibles; (B) Infectious; (C) Time evolution of both R_{eff} and $\beta(t)$. In (A) and (B) the black lines are the true values, the blue lines are the median of the posterior, the mauve areas are the 50% CI and the light blue areas the 95% CI. In (C) the black line is the true values of R_{eff} , the blue line is the median of the posterior, and the dashed lines the 95% CI of R_{eff} , the red dot line is the true time evolution of $\beta(t)$ and the red line the median of its posterior. Model parameters as in Fig 1 and S6 Fig.

(PDF)

S6 Fig. Prior and posterior distributions for the SIRS model inferences of S4 Fig. $I(0)$, $S(0)$ initial values, $1/\alpha$ is the average duration of immunity, γ is the recovery rate and σ is the volatility of the Brownian process of $\beta(t)$. The blue distributions are the priors and the discrete histograms are the posteriors. The medians of the prior distributions for $I(0)$, $S(0)$, $1/\alpha$ and $1/\gamma$, are the “true values” used for the simulations of the observed incidences.

(PDF)

S7 Fig. Reconstruction of both the incidence (A) and the time evolution $\beta(t)$ (B) for the SIRS model as in Fig 1 but σ the volatility of the Brownian process of $\beta(t)$ is the only parameter inferred. Model parameters as in Fig 1 and S8 Fig.

(PDF)

S8 Fig. The trace of the MCMC chain and the prior and posterior distributions for the SIRS model inferences of [S7 Fig](#) when σ is the volatility of the Brownian process of $\beta(t)$ is the only parameter inferred.

(PDF)

S9 Fig. Reconstruction of both the incidence (A) and the time evolution $\beta(t)$ (B) for the SIRS model as in [S4 Fig](#) but the logarithm transformation of the Brownian process of $\beta(t)$ is not used: $d\theta(t) = \sigma dB(t)$. Model parameters as in [Fig 1](#) and [S4 Fig](#).

(PDF)

S10 Fig. As in [S6 Fig](#) but the logarithm transformation of the Brownian process of $\beta(t)$ is not used: $d\theta(t) = \sigma dB(t)$.

(PDF)

S11 Fig. Reconstruction of both the incidence (A) and the time evolution $\beta(t)$ (B) with the true SIRS model. In (A) the black points are observations generated with a Poisson process with a mean equal to the incidence simulated by the model. In (B) the black points are the true values of $\beta(t) = \beta_0(1 + \beta_1 \sin(2\pi t/365 + 2\pi\phi))$. The blue lines are the median of the posterior, the mauve areas are the 50% Credible Intervals (CI) and the light blue areas the 95% CI. For all the figures, the observation process is also applied to the inferred incidence trajectory. The time unit of the model is day, the initial date is arbitrary (2000-01-09) and parameters used for the SIRS model are as follows: $\mu = 1/(50*365)$, $\alpha = 1/(7*365)$, $\gamma = 1/14$, $\beta_0 = 0.65$, $\beta_1 = 0.4$, $\phi = -0.2$, $\rho = 1$, $N = 10000$, $S(0) = 600$, $I(0) = 30$. The prior and posterior distributions of the inferred parameters are in [S13 Fig](#)

(PDF)

S12 Fig. Simulation of the true SIRS model: (A) Susceptibles; (B) Infectious; (C) Time evolution of both R_{eff} and $\beta(t)$. In (A) and (B) the black lines are the true values, the blue lines are the median of the posterior, the mauve areas are the 50% CI and the light blue areas the 95% CI. In (C) the black line is the true values of R_{eff} , the blue line is the median of the posterior, and the dashed lines the 95% CI of R_{eff} ; the red dot line is the true time evolution of $\beta(t)$ and the red line the median of its posterior. Model parameters as in [S11 Fig](#) and [S13 Fig](#).

(PDF)

S13 Fig. Prior and posterior distributions for the true SIRS model inferences of [S11 Fig](#). $I(0)$, $S(0)$ initial values, β_0 , β_1 , ϕ , the parameters of the sinusoidal β , $1/\alpha$ is the average duration of immunity, γ is the recovery rate and ρ is the reporting rate. The blue distributions are the priors and the discrete histograms are the posteriors. The medians of the prior distributions are the “true values” used for the simulations of the observed incidences.

(PDF)

S14 Fig. Simulation of the SIRS model: (A) Susceptibles; (B) Infectious; (C) Time evolution of both R_{eff} and $\varepsilon_S(t)$. In (A) and (B) the black lines are the true values, the blue lines are the median of the posterior, the mauve areas are the 50% CI and the light blue areas the 95% CI. In (C) the black line is the true values of R_{eff} , the blue line is the median of the posterior and the dashed lines the 95% CI of R_{eff} ; the red dot line is the true time evolution of $\varepsilon_S(t)$ and the red line the median of its posterior. Model parameters as in [Fig 4](#) and [S15 Fig](#).

(PDF)

S15 Fig. Prior and posterior distributions for the SIRS model inferences of [Fig 4](#). $I(0)$, $S(0)$ initial values, $1/\alpha$ is the average duration of immunity, γ is the recovery rate and σ is the volatility of the Brownian process of $\varepsilon_S(t)$. The blue distributions are the priors and the discrete

histograms are the posteriors. The medians of the prior distributions for $I(0)$, $S(0)$, $1/\alpha$ and $1/\gamma$, are the “true values” used for the simulations of the observed incidences.
 (PDF)

S16 Fig. Prior and posterior distributions for the parameters of the SIRS flu model. $I(0)$, $S(0)$ initial values expressed in percentage of the population N , $\beta(0)$ initial value of $\beta(t)$, i imported infectious, $1/\alpha$ is the average duration of immunity, γ is the recovery rate and σ is the volatility of the Brownian process of $\beta(t)$. The blue distributions are the priors and the discrete histograms are the posteriors. Prior values are adapted from [46].
 (PDF)

S17 Fig. Prior and posterior distributions for the parameters of the SEIR dengue model. $E(0)$, $S(0)$ expressed in percentage of the population N , $\beta(0)$ the initial value of $\beta(t)$, γ is the recovery rate, i imported infectious, $1/\alpha$ is the average duration of immunity, σ is the volatility of the Brownian process of $\beta(t)$ and ρ the reporting rate. The blue distributions are the priors and the discrete histograms are the posteriors. Prior values are adapted from [63].
 (PDF)

S18 Fig. Association between dengue transmission rate and monthly average maximum temperature recorded at the Phnom Penh International Airport (Cambodia). (A) Time evolution of the normalized $\beta(t)$ (blue line) and normalized average temperature (red line). (B) and (C) Wavelet Power Spectrum (WPS) [48,70] of the two time series. The graph on the right shows the average WPS. (D) Wavelet coherence [48,70] between the reconstructed $\beta(t)$ and average temperature. In (B), (C) and (D) the colors code for low values in white to high values in dark red. The dashed lines show the 95% CI computed with adapted bootstrappes [71], in (C) the 90% and the 95% CI have been plotted. (E) The evolution of the phase of the two time series computed based on wavelet decomposition for the seasonal mode, blue dashed line for the normalized $\beta(t)$ red dashed line for the normalized averaged temperature and black dotted line for their phase difference. The graph on the right shows the distribution of the phase differences.
 (PDF)

S19 Fig. Association between dengue transmission rate and monthly average minimum temperature recorded at the Phnom Penh International Airport (Cambodia). (A) Time evolution of the normalized $\beta(t)$ (blue line) and normalized average temperature (red line). (B) and (C) Wavelet Power Spectrum (WPS) [48,70] of the two time series. The graph on the right shows the average WPS. (D) Wavelet coherence [48,70] between the reconstructed $\beta(t)$ and average temperature. In (B), (C) and (D) the colors code for low values in white to high values in dark red. The dashed lines show the 95% CI computed with adapted bootstrappes [71], in (C) the 90% and the 95% CI have been plotted. (E) The evolution of the phase of the two time series computed based on wavelet decomposition for the seasonal mode, blue dashed line for the normalized $\beta(t)$ red dashed line for the normalized averaged temperature and black dotted line for their phase difference. The graph on the right shows the distribution of the phase differences.
 (PDF)

S20 Fig. Association between dengue transmission rate and monthly rainfall recorded at the Phnom Penh International Airport (Cambodia). (A) Time evolution of the normalized $\beta(t)$ (blue line) and normalized monthly rainfall (red line). (B) and (C) Wavelet Power Spectrum (WPS) [48,70] of the two time series. The graph on the right shows the average WPS. (D) Wavelet coherence [48,70] between the reconstructed $\beta(t)$ and monthly rainfall. In (B), (C)

and (D) the colors code for low values in white to high values in dark red. The dashed lines show the 95% CI computed with adapted bootstrappes [71], in (C) the 90% and the 95% CI have been plotted. (E) The evolution of the phase of the two time series computed based on wavelet decomposition for the seasonal mode, blue dashed line for the normalized $\beta(t)$ red dashed line for the normalized monthly rainfall and black dotted line for their phase difference. The graph on the right shows the distribution of the phase differences.
(PDF)

S21 Fig. Association between dengue transmission rate and the Dipole Mode Index (DMI), a proxy of Ocean Indian Dipole (see [72] and <http://www.jamstec.go.jp/frcgc/research/d1/iod/HTML/Dipole%20Mode%20Index.html>). (A) Time evolution of the normalized $\beta(t)$ (blue line) and normalized DMI (red line). (B) and (C) Wavelet Power Spectrum (WPS) [48,70] of the two time series. The graph on the right shows the average WPS. (D) Wavelet coherence [48,70] between the reconstructed $\beta(t)$ and DMI. In (B), (C) and (D) the colors code for low values in white to high values in dark red. The dashed lines show the 95% CI computed with adapted bootstrappes [71], in (C) the 90% and the 95% CI have been plotted. (E) The evolution of the phase of the two time series computed based on wavelet decomposition for the seasonal mode, blue dashed line for the normalized $\beta(t)$ red dashed line for the normalized DMI and black dotted line for their phase difference. The graph on the right shows the distribution of the phase differences.

(PDF)

Acknowledgments

BC would like to express his gratitude to the “Institut Ecologie et Environnement” of CNRS for facilitating his research activity thanks to a “Délégation CNRS” for 2 years in the URA 3012 “Hosts, Vectors and Infectious Agents”. “Lovely thanks” to Una Ni Mhaoldhomhnaigh who has smoothed our “French English”.

Author Contributions

Conceptualization: Bernard Cazelles.

Data curation: Bernard Cazelles.

Formal analysis: Bernard Cazelles, Clara Champagne.

Funding acquisition: Bernard Cazelles.

Investigation: Bernard Cazelles, Clara Champagne.

Methodology: Bernard Cazelles, Clara Champagne, Joseph Dureau.

Project administration: Bernard Cazelles.

Resources: Bernard Cazelles.

Software: Joseph Dureau.

Supervision: Bernard Cazelles.

Validation: Bernard Cazelles.

Visualization: Bernard Cazelles.

Writing – original draft: Bernard Cazelles, Clara Champagne, Joseph Dureau.

Writing – review & editing: Bernard Cazelles, Clara Champagne, Joseph Dureau.

References

1. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. Global trends in emerging infectious diseases. *Nature*. 2008; 451: 990–994. <https://doi.org/10.1038/nature06536> PMID: 18288193
2. Cazelles B, Hales S. Infectious diseases, climate influences, and nonstationarity. *PLoS Med*. 2006; 3(8): e328. <https://doi.org/10.1371/journal.pmed.0030328> PMID: 16903777
3. Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P. Seasonality and the dynamics of infectious diseases. *Ecology letters*. 2006; 9(4): 467–484. <https://doi.org/10.1111/j.1461-0248.2005.00879.x> PMID: 16623732
4. Fisman DN. Seasonality of infectious diseases. *Annu. Rev. Public Health*. 2007; 28: 127–143. <https://doi.org/10.1146/annurev.publhealth.28.021406.144128> PMID: 17222079
5. Cazelles B, Chavez M, McMichael AJ, Hales S. Nonstationary influence of El Nino on the synchronous dengue epidemics in Thailand. *PLoS Med*. 2005; 2(4): e106. <https://doi.org/10.1371/journal.pmed.0020106> PMID: 15839751
6. Constantin de Magny GC, Cazelles B, Guégan JF. Cholera threat to humans in Ghana is influenced by both global and regional climatic variability. *EcoHealth*. 2006; 3(4): 223–231.
7. Laneri K, Bhadra A, Ionides EL, Bouma M, Dhiman RC, Yadav RS, Pascual M. 2010 Forcing versus feedback: epidemic malaria and monsoon rains in northwest India. *PLoS Comput. Biol*. 2010; 6: e1000898. <https://doi.org/10.1371/journal.pcbi.1000898> PMID: 20824122
8. Metcalf CJE, Walter KS, Wesolowski A, Buckee CO, Shevliakova E, Tatem AJ, Boos WR, Weinberger DM, Pitzer VE. Identifying climate drivers of infectious disease dynamics: Recent advances and challenges ahead. *Proc. R. Soc. B*. 2017; 284: 20170901. <https://doi.org/10.1098/rspb.2017.0901> PMID: 28814655
9. Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, Nicoll A. Closure of schools during an influenza pandemic. *The Lancet Infectious Diseases*. 2009; 9(8): 473–481. [https://doi.org/10.1016/S1473-3099\(09\)70176-8](https://doi.org/10.1016/S1473-3099(09)70176-8) PMID: 19628172
10. Wu JT, Cowling BJ, Lau EH, Ip DK, Ho LM, Tsang T, Chuang SK, Leung PY, Lo SV, Liu SH, Riley S. School closure and mitigation of pandemic (H1N1) 2009, Hong Kong. *Emerging Infectious Diseases*. 2010; 16(3): 538–541. <https://doi.org/10.3201/eid1603.091216> PMID: 20202441
11. Ewing A, Lee EC, Viboud C, Bansal S. Contact, travel, and transmission: The impact of winter holidays on influenza dynamics in the United States. *The Journal of Infectious Diseases*. 2016; 215(5): 732–739. <https://doi.org/10.1093/infdis/jiw642>
12. Khan K, Memish ZA, Chabria A, Liauw J, Hu W, Janes DA, Sears J, Arino J, Macdonald M., Calderon F., Raposo P, Heidebrecht C, Wang J, Chan A, Brownstein J, Gardam M. Global public health implications of a mass gathering in Mecca, Saudi Arabia during the midst of an influenza pandemic. *Journal of Travel Medicine*. 2010; 17(2): 75–81. <https://doi.org/10.1111/j.1708-8305.2010.00397.x> PMID: 20412172
13. Ferrari MJ, Djibo A, Grais RF, Bharti N, Grenfell BT, Bjørnstad ON. Rural–urban gradient in seasonal forcing of measles transmission in Niger. *Proc. R. Soc. B*. 2010; 277: 2775–2782. <https://doi.org/10.1098/rspb.2010.0536> PMID: 20427338
14. Funk S, Salathé M, Jansen VA. Modelling the influence of human behaviour on the spread of infectious diseases: A review. *J. R. Soc. Interface*. 2010; 7: 1247–1256. <https://doi.org/10.1098/rsif.2010.0142> PMID: 20504800
15. Verelst F, Willem L, Beutels P. Behavioural change models for infectious disease transmission: A systematic review (2010–2015). *J. R. Soc. Interface*. 2016; 13: 20160820. <https://doi.org/10.1098/rsif.2016.0820> PMID: 28003528
16. Kucharski AJ, Kwok KO, Wei VW, Cowling BJ, Read JM, Lessler J, Cummings DA, Riley S. The contribution of social behavior to the transmission of influenza A in a human population. *PLoS Pathogen*. 2014; 10: e1004206. <https://doi.org/10.1371/journal.ppat.1004206>
17. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2008.
18. 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; 347(6227), aaa4339. <https://doi.org/10.1126/science.aaa4339> PMID: 25766240
19. Metcalf CJE, Lessler J. Opportunities and challenges in modeling emerging infectious diseases. *Science*. 2017; 357(6347), 149–152. <https://doi.org/10.1126/science.aam8335> PMID: 28706037
20. Axelsen JB, Yaari R, Grenfell BT, Stone L. Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proc Natl Acad Sci USA*. 2014; 111(26): 9538–9542. <https://doi.org/10.1073/pnas.1321656111> PMID: 24979763

21. Dorigatti I, Cauchemez S, Ferguson NM. Increased transmissibility explains the third wave of infection by the 2009 H1N1 pandemic virus in England. *Proc Natl Acad Sci USA*. 2013; 110(33): 13422–13427. <https://doi.org/10.1073/pnas.1303117110> PMID: 23882078
22. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*. 2014; 160(6): 509–516. <https://doi.org/10.1093/aje/kwh255>
23. Cauchemez S, Boëlle PY, Donnelly CA, Ferguson NM, Thomas G, Leung GM, Hedley AJ, Anderson R., Valleron AJ. Real-time estimates in early detection of SARS. *Emerging Infectious Diseases*. 2006; 12(1): 110–113. <https://doi.org/10.3201/eid1201.050593> PMID: 16494726
24. Nishiura H, Chowell G, Heesterbeek H, Wallinga J. The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course. *J. R. Soc. Interface*. 2009; 7: 297–307. <https://doi.org/10.1098/rsif.2009.0153> PMID: 19570792
25. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*. 2013; 178(9), 1505–1512. <https://doi.org/10.1093/aje/kwt133> PMID: 24043437
26. Ionides EL, Bretó C, King AA. Inference for nonlinear dynamical systems. *Proc Natl Acad Sci USA*. 2006; 103(49): 18438–18443. <https://doi.org/10.1073/pnas.0603181103> PMID: 17121996
27. Churcher TS, Cohen JM, Novotny J, Ntshalintshali N, Kunene S, Cauchemez S. Measuring the path toward malaria elimination. *Science*. 2014; 344(6189): 1230–1232. <https://doi.org/10.1126/science.1251449> PMID: 24926005
28. Coelho FC, De Carvalho LM. Estimating the attack ratio of dengue epidemics under time-varying force of infection using aggregated notification data. *Scientific report*. 2015; 5: 18455. <http://dx.doi.org/10.1038/srep18455>
29. Ellner SP, Bailey BA, Bobashev GV, Gallant AR, Grenfell B T, Nychka DW. Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. *The American Naturalist* 1998; 151: 425–440. <https://doi.org/10.1086/286130> PMID: 18811317
30. King AA, Ionides EL, Pascual M, Bouma MJ. Inapparent infections and cholera dynamics. *Nature*. 2008; 454: 877–880. <https://doi.org/10.1038/nature07084> PMID: 18704085
31. Bhadra A, Ionides EL, Laneri K, Pascual M, Bouma M, Dhiman RC (2011). Malaria in Northwest India: Data analysis via partially observed stochastic differential equation models driven by Lévy noise. *Journal of the American Statistical Association*. 2011; 106: 440–451. <https://doi.org/10.1198/jasa.2011.ap10323>
32. He D, Dushoff J, Day T, Ma J, Earn DJ. Mechanistic modelling of the three waves of the 1918 influenza pandemic. *Theoretical Ecology*. 2011; 4: 283–288. <https://doi.org/10.1007/s12080-011-0123-3>
33. Martinez-Bakker M, King AA, Rohani P. Unraveling the transmission ecology of polio. *PLoS Biol*. 2015; 13: e1002172. <https://doi.org/10.1371/journal.pbio.1002172> PMID: 26090784
34. Cazelles B, Chau NP. Adaptive dynamic modelling of HIV/AIDS epidemics using extended Kalman filter. *Journal of Biological Systems*. 1995; 3: 759–768. <https://doi.org/10.1142/S0218339095000691>
35. Cazelles B, Chau, NP. Using the Kalman filter and dynamic models to assess the changing HIV/AIDS epidemic. *Mathematical Biosciences*. 1997; 140(2): 131–154. [https://doi.org/10.1016/S0025-5564\(96\)00155-1](https://doi.org/10.1016/S0025-5564(96)00155-1) PMID: 9046772
36. Dureau J, Kalogeropoulos K, Baguelin M. Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics*. 2013; 14(3): 541–555. <https://doi.org/10.1093/biostatistics/kxs052> PMID: 23292757
37. Angulo MT, Velasco-Hernandez JX. Robust qualitative estimation of time-varying contact rates in uncertain epidemics. *Epidemics*. 2018; <https://doi.org/10.1016/j.epidem.2018.03.001> PMID: 29567063
38. Smirnova A, Chowell G, de Camp L. Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the SEIR model. *Bulletin of Mathematical Biology*. 2017; on-line <https://doi.org/10.1007/s11538-017-0284-3> PMID: 28466232
39. Finkenstädt BF, Grenfell BT. Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C*. 2000; 49(2): 187–205.
40. Bjørnstad ON, Finkenstädt BF, Grenfell BT. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*. 2002; 72(2): 169–184.
41. Cobelli C, DiStefano JJ. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *Am. J. Physiol.—Regul. Integr. Comp. Physiol.* 1980; 239: R7–R24.
42. Miao H, Xia X, Perelson AS, Wu H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM review* 2011; 53: 3–39. <https://doi.org/10.1137/090757009> PMID: 21785515
43. Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*. 2010; 72: 269–342.

44. Hethcote HW. The mathematics of infectious diseases. SIAM review. 2000; 42(4): 599–653.
45. Lee EC, Kelly MR, Ochocki BM, Akinwumi SM, Hamre KES, Tien JH, Eisenberg MC. Model distinguishability and inference robustness in mechanisms of cholera transmission and loss of immunity. Journal of Theoretical Biology. 2017; 420: 68–81. <https://doi.org/10.1016/j.jtbi.2017.01.032> PMID: 28130096
46. Yaari R, Katriel G, Huppert A, Axelsen JB, Stone L. Modelling seasonal influenza: the role of weather and punctuated antigenic drift. J R Soc Interface. 2013; 10: 20130298. <https://doi.org/10.1098/rsif.2013.0298> PMID: 23676899
47. Teurlai M, Huy R, Cazelles B, Duboz R, Baehr C, Vong S. Can Human Movements Explain Heterogeneous Propagation of Dengue Fever in Cambodia? PLoS Negl Trop Dis. 2012; 6(12): e1957. <https://doi.org/10.1371/journal.pntd.0001957> PMID: 23236536
48. Cazelles B, Chavez M, Constantin de Magny G, Guégan JF, Hales S. Time-dependent spectral analysis of epidemiological time-series with wavelets. J. R. Soc. Interface. 2007; 4: 625–636. <https://doi.org/10.1098/rsif.2007.0212> PMID: 17301013
49. Martínez-Bello DA, López-Quílez A, Torres-Prieto A. Bayesian dynamic modeling of time series of dengue disease case counts. PLoS Negl Trop Dis. 2017; 11(7): e0005696. <https://doi.org/10.1371/journal.pntd.0005696> PMID: 28671941
50. Opatowski L, Baguelin M, Eggo RM. Influenza interaction with cocirculating pathogens and its impact on surveillance, pathogenesis, and epidemic profile: A key role for mathematical modelling. PLoS Pathogen. 2018; 14: e1006770. <https://doi.org/10.1371/journal.ppat.1006770>
51. Funk S, Camacho A, Kucharski AJ, Eggo RM, Edmunds WJ. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. Epidemics. 2018; 22: 56–61. <https://doi.org/10.1016/j.epidem.2016.11.003> PMID: 28038870
52. Camacho A, Kucharski A, Aki-Sawyerr Y, White MA, Flasche S, Baguelin M, Pollington T, Carney JR, Glover R, Smout E, Tiffany A, Edmunds WJ, Funk. Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. PLoS Curr. 2015; 7 (<http://dx.doi.org/10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2>).
53. Jacquez JA, Greif P. Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. Mathematical Biosciences. 1985; 77: 201–227.
54. Tunali E, Tarn TJ. New results for identifiability of nonlinear systems. IEEE Transactions on Automatic Control. 1987; 32: 146–154.
55. Audoly S, Bellu G, D'Angio L, Saccomani MP, Cobelli C. Global identifiability of nonlinear models of biological systems. IEEE Transactions on Biomedical Engineering. 2001; 48: 55–65. <https://doi.org/10.1109/10.900248> PMID: 11235592
56. Evans ND, White LJ, Chapman MJ, Godfrey KR, Chappell MJ. The structural identifiability of the susceptible infected recovered model with seasonal forcing. Mathematical Biosciences. 2005; 194: 175–197. <https://doi.org/10.1016/j.mbs.2004.10.011> PMID: 15854675
57. Chapman JD, Evans ND. The structural identifiability of susceptible–infective–recovered type epidemic models with incomplete immunity and birth targeted vaccination. Biomedical Signal Processing and Control. 2009; 4: 278–284.
58. Champagne C, Salthouse DG, Paul R, Cao-Lormeau VM, Roche B, Cazelles B. Structure in the variability of the basic reproductive number (R_0) for Zika epidemics in the Pacific islands. eLife. 2016; 5: e19874. <https://doi.org/10.7554/eLife.19874> PMID: 27897973
59. Bengtsson T, Bickel P, and Bo Li B. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In Probability and Statistics: Essays in Honor of David A. Freedman, Vol. 2; 2008. pp. 316–334.
60. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface. 2007; 6: 187–202. <https://doi.org/10.1098/rsif.2008.0172>
61. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian Computation. PLoS Comput Biol. 2013; 9: e1002803. <https://doi.org/10.1371/journal.pcbi.1002803> PMID: 23341757
62. Metcalf CJE, Farrar J, Cutts FT, Basta NE, Graham AL, Lessler J, Ferguson NM, Burke DS, Grenfell BT. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. The Lancet. 2016; 388(10045): 728–730. [https://doi.org/10.1016/S0140-6736\(16\)30164-7](https://doi.org/10.1016/S0140-6736(16)30164-7)
63. Champagne C, Paul R, Ly S, Duong V, Leang R, Cazelles B. Dengue modeling in rural Cambodia: statistical performance versus epidemiological relevance. bioRxiv. 2017; 208876.
64. Bretó C, He D, Ionides EL, King AA. Time Series Analysis via Mechanistic Models. The Annals of Applied Statistics. 2009; 3: 319–348.

65. Dureau J, Ballesteros S, Bogich T. SSM: Inference for time series analysis with State Space Models. 2013; <https://github.com/JDureau/ssm/blob/master/doc/doc.pdf>.
66. Plummer M, Best N, Cowles K, Vines K. Coda: convergence diagnosis and output analysis for MCMC. *R news.* 2006; 6: 7–11.
67. Geweke J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian Statistics 4.* Clarendon Press, Oxford, UK; 1992.
68. Heidelberger P, Welch PD. Simulation run length control in the presence of an initial transient. *Opsns Res.* 1983; 31: 1109–11044.
69. Torrence C, Compo GP. A practical guide to wavelet analysis. *Bull Am Meteorol Soc.* 1998; 79: 61–78.
70. Cazelles B, Chavez M, Berteaux D, Ménard F, Vik JO, Jenouvrier S, Stenseth NC. Wavelet analysis of ecological time series. *Oecologia.* 2008; 156: 287–304. <https://doi.org/10.1007/s00442-008-0993-2> PMID: 18322705
71. Cazelles B, Cazelles K, Chavez M. Wavelet analysis in ecology and epidemiology: impact of statistical tests. *J. R. Soc. Interface.* 2014; 11: 20130585. <https://doi.org/10.1098/rsif.2013.0585> PMID: 24284892
72. Saji NH, Goswami BN, Vinayachandran PN, Yamagata T. A dipole mode in the tropical Indian Ocean. *Nature.* 1999; 401(6751): 360–363. <https://doi.org/10.1038/43854> PMID: 16862108