Structure in the variability of the basic reproductive number (R_0) for Zika epidemics in the Pacific islands

Clara Champagne *[1,2], David Georges Salthouse[1], Richard Paul[3,4], Van-Mai Cao-Lormeau[5], Benjamin Roche[6] and Bernard Cazelles †[1,6]

[1]IBENS, UMR 8197 CNRS-ENS Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris, France
[2]Centre de Recherche en Economie et Statistiques (CREST), 15 boulevard Gabriel Péri, 92245 Malakoff cedex, 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]Unit of Emerging Infectious Diseases, Institut Louis Malardé, 98713 Papeete, Tahiti, French Polynesia
[6]International Center for Mathematical and Computational Modeling of Complex Systems (UMMISCO), UMI 209 UPMC/IRD, Bondy cedex, France

*champagn@biologie.ens.fr
†cazelles@biologie.ens.fr

Abstract

Before the outbreak that reached the Americas in 2015, Zika virus (ZIKV) circulated in Asia and the Pacific: these past epidemics can be highly informative on the key parameters driving virus transmission, such as the basic reproduction number (R_0). We compare two compartmental models with different mosquito representations, using surveillance and seroprevalence data for several ZIKV outbreaks in Pacific islands (Yap, Micronesia 2007, Tahiti and Moorea, French Polynesia 2013-2014, New Caledonia 2014). Models are estimated in a stochastic framework with state-of-the-art Bayesian techniques. R_0 for the Pacific ZIKV epidemics is estimated between 1.5 and 4.1, the smallest islands displaying higher and more variable values. This relatively low range of R_0 suggests that intervention strategies developed for other flaviviruses should enable as, if not more effective control of ZIKV. Our study also highlights the importance of seroprevalence data for precise quantitative analysis of pathogen propagation, to design prevention and control strategies.

In May 2015, the first local cases of Zika were recorded in Brazil and by December of the same year the number of cases had surpassed 1.5 million. On February 2016, the World Health Organization declared Zika as a public health emergency of international concern [1] and in March 2016, local transmission of Zika was recognized in 34 countries. Previously the Zika virus had circulated in Africa and Asia but only sporadic human cases had been reported. In 2007 the outbreak on Yap (Micronesia) was the first Zika outbreak outside Africa and Asia.[2] Since, Zika outbreaks have been also reported in French Polynesia and in New Caledonia[3;4] between 2013 and 2014 and subsequently, there have been cases of Zika disease in the Cook Islands, the Solomon Islands, Samoa, Vanuatu, and Easter Island (Chile) (see Fig. 1 in reference[5]).

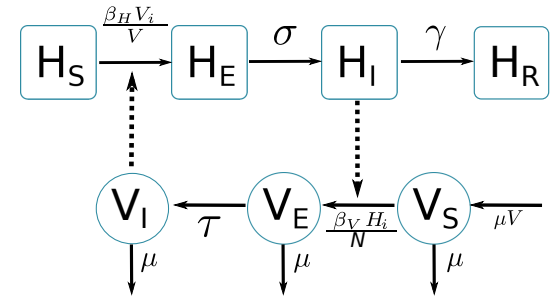
Zika virus (ZIKV) is a flavivirus, mostly transmitted via the bites of infected *Aedes* mosquitoes, although non-vector-borne transmission has been documented (sexual and maternofoetal transmission, laboratory contamination, transmission through transfusion).[6] The most common clinical manifestations include rash, fever, arthralgia, and conjunctivitis[6] but a large proportion of infections are asymptomatic or trigger mild symptoms that can remain unnoticed. Nevertheless, the virus may be involved in many severe neurological complications, including Guillain-Barre syndrome[7] and microcephaly in newborns.[8] These complications and the impressive speed of its geographically propagation make the Zika pandemic a public health threat.[1] This reinforces the urgent need to characterize the different facets of virus transmission and to evaluate its dispersal capacity. We address this here by estimating the key parameters of ZIKV transmission, including the basic reproduction number (R_0), based on previous epidemics in the Pacific islands.

Defined as the average number of secondary cases caused by one typical infected individual in an entirely susceptible population, the basic reproduction number (R_0) is a central parameter in epidemiology used to quantify the magnitude of ongoing outbreaks and

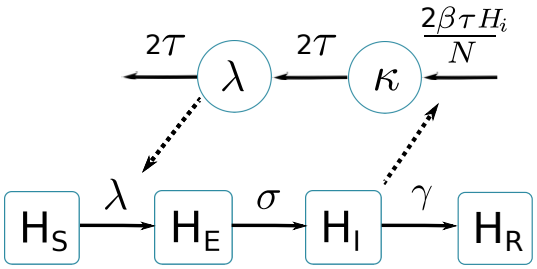
it provides insight when designing control interventions.[9] It is nevertheless complex to estimate[9;10], and therefore, care must be taken when extrapolating the results obtained in a specific setting, using a specific mathematical model. In the present study, we explore the variability of R_0 using two state-of-the-art models in several settings that had Zika epidemics in different years and that vary in population size (Yap, Micronesia 2007, Tahiti and Moorea, French Polynesia 2013-2014, and New Caledonia 2014). These three countries were successively affected by the virus, resulting in the first significant human outbreaks and they differ in several ways, including population size and location specific features. Hence, the comparison of their parameter estimates can be highly informative on the intrinsic variability of R_0. For each setting, we compare two compartmental models using different parameters defining the mosquito populations. Both models are considered in a stochastic framework, a necessary layer of complexity given the small population size and state-of-the-art Bayesian inference techniques[11] are used for parameter estimation.

Results

We use mathematical transmission models and data from surveillance systems and seroprevalence surveys for several ZIKV outbreaks in Pacific islands (Yap, Micronesia 2007[2], Tahiti and Moorea, French Polynesia 2013-2014[12–14], New Caledonia 2014[15]) to quantify the ZIKV transmission variability.



(a)



(b)

Figure 1 – **Graphical representation of compartmental models.** 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. a) Pandey model.[16] 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 infectious class H_I; infectious individuals (H_I) then recover at rate γ; V_S susceptible vectors; V_E infected (not yet infectious) vectors; V_I infectious vectors; V constant size of total mosquito population; τ is the rate at which V_E-vectors move to infectious class V_I; vectors die at rate μ. b) Laneri model.[17] 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 infectious class H_I; infectious individuals (H_I) then recover at rate γ; implicit vector-borne transmission is modeled 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; τ is the transition rate associated to the extrinsic incubation period.

Two compartmental models with vector-borne transmission are compared (cf. Figure 1). Both models use a Susceptible-Exposed-Infected-Resistant (SEIR) framework to describe the virus transmission in the human population, but differ in their representation of the mosquito population. Figure 1.a. is a schematic representation derived from Pandey et al.[16]

and formulates explicitly the mosquito population, with a Susceptible-Exposed-Infected (SEI) dynamic to account for the extrinsic incubation period (time taken for viral dissemination within the mosquito).

By contrast, in the second model (Figure 1.b.) based on Laneri et al.[17] the vector is modeled implicitly: the two compartments κ and λ do not represent the mosquito population but the force of infection for vector to human transmission. This force of infection passes through two successive stages in order to include the delay associated with the extrinsic incubation period: κ stands for this latent phase of the force of infection whereas λ corresponds directly to the rate at which susceptible humans become infected.

The basic reproduction number of the models (R_0) is calculated using the next Generation Matrix method:[9]

$$R_0^{Pandey} = \sqrt{\frac{\beta_H \beta_V \tau}{\gamma \mu (\mu + \tau)}}$$

$$R_0^{Laneri} = \sqrt{\frac{\beta}{\gamma}}$$

In addition, we consider that only a fraction ρ of the total population is involved in the epidemic, due to spatial heterogeneity, immuno-resistance, or cross-immunity. For both models we define $N = \rho \cdot H$ with H the total size of the population reported by census.

The dynamics of ZIKV transmission in these islands is highly influenced by several sources of uncertainties. In particular, the small population size (less than 7,000 inhabitants in Yap) leads to high variability in transmission rates. Therefore all these models are simulated in a discrete stochastic framework (Poisson with stochastic rates[18]), to take this phenomenon into account. Stochasticity requires specific inference techniques : thus estimations are performed with PMCMC algorithm (particle Markov Chain Monte Carlo[11]).

Using declared Zika cases from different settings, the two stochastic models (Fig. 1) were fitted (Figs 2-3). These models allow us to describe the course of the observed number of cases and estimate the number of secondary cases generated, R_0. Our estimates of R_0 lie between 1.6 (1.5-1.7) and 3.2 (2.4-4.1) and vary notably with respect to settings and models (Figures 2-3 and Tables 1-2). Strikingly, Yap displays consistently higher values of R_0 in both models and in general, there is an inverse relationship between island size and both the value and variability of R_0. This phenomenon may be explained by the higher stochasticity and extinction probability associated with smaller populations and can also reflect the information contained in the available data. However, the two highly connected islands in French Polynesia, Tahiti and Moorea, display similar values despite their differing sizes.

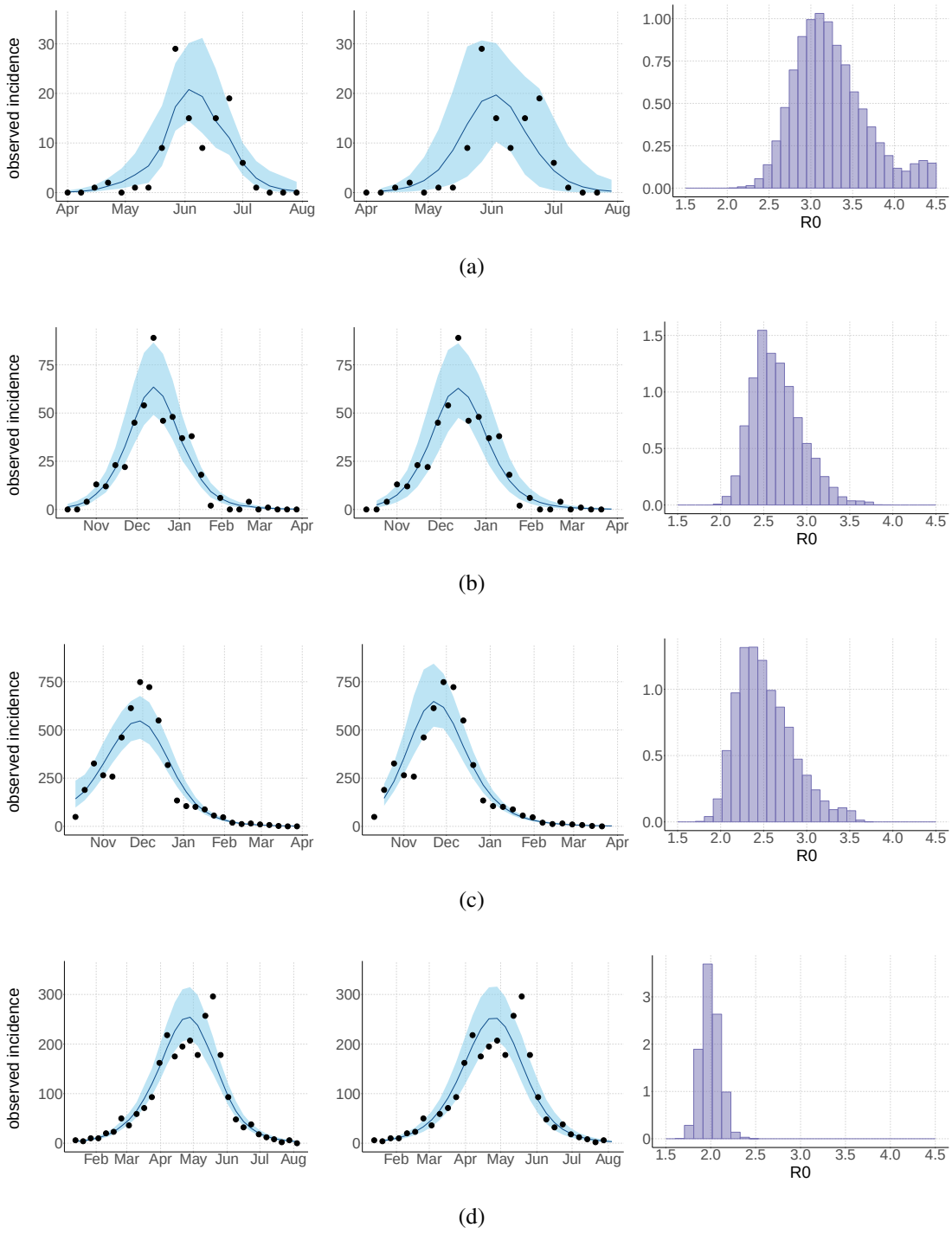


Figure 2 – **Results using the Pandey model.** Posterior median number of observed Zika cases (solid line), 95% credible intervals (shaded blue area) and data points (black dots). First column: particle filter fit. Second column: Simulations from the posterior density. Third column: R_0 posterior distribution. a) Yap. b) Moorea. c) Tahiti. d) New Caledonia. The estimated seroprevalences at the end of the epidemic (with 95% credibility intervals) are: a) 73% (CI95: 68-77, observed 73%); b) 49% (CI95: 45-53, observed 49%); c) 49% (CI95: 45-53, observed 49%); d) 39% (CI95: 8-92). See Figure 4.

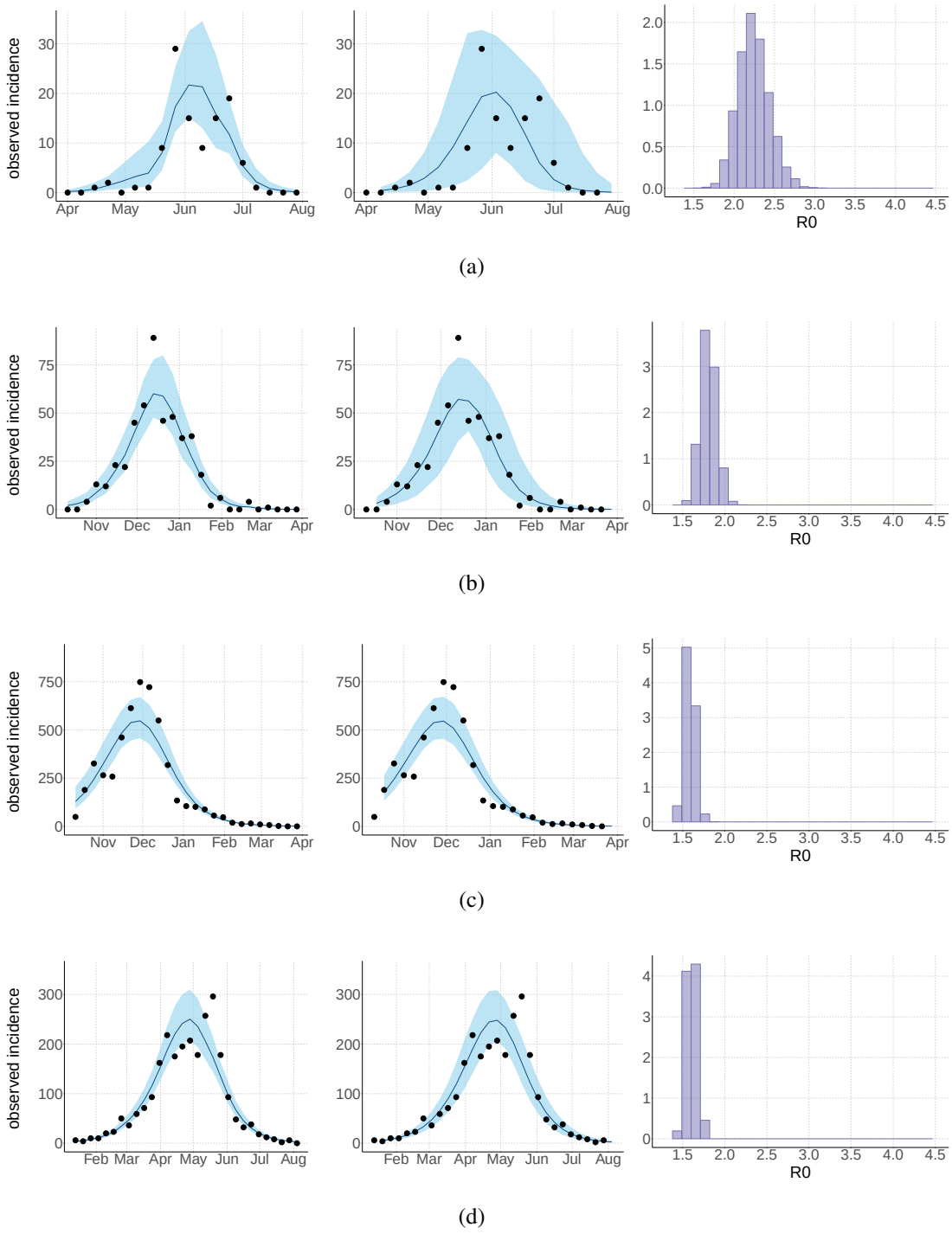


Figure 3 – **Results using the Laneri model.** Posterior median number of observed Zika cases (solid line), 95% credible intervals (shaded blue area) and data points (black dots). First column: particle filter fit. Second column: Simulations from the posterior density. Third column: R_0 posterior distribution. a) Yap. b) Moorea. c) Tahiti. d) New Caledonia. The estimated seroprevalences at the end of the epidemic (with 95% credibility intervals) are: a) 72% (CI95: 68-77, observed 73%); b) 49% (CI95: 45-53, observed 49%); c) 49% (CI95: 45-53, observed 49%); d) 65% (CI95: 24-91). See Figure 5.

Regarding model variability, R_0 estimates are always higher and coarser with the Pandey model than with the Laneri model (cf. Tables 1-2). The Pandey model has two additional estimated parameters (in particular, the mosquito lifespan), which can explain the higher variability of the output. It is worth noting that these parameters are very sensitive (see Materials and methods). The difference in R_0 may also be linked to the difference in the estimated initial number of infected individuals (H_{I0}), which is higher in the Laneri model than in the Pandey model. Because of the high proportion of asymptomatic cases (the ratio of asymptomatic:symptomatic is estimated to be 1:1.3, V.-M Cao-Lormeau personal communication), it is hard to determine which scenario is more realistic, the time between introduction of the disease into the island and the first reported symptomatic case being unknown in most settings.

PANDEY MODEL		Yap	Moorea	Tahiti	New Caledonia
Population size	H	6,892	16,200	178,100	268,767
Basic reproduction number	R_0	3.2 (2.4 – 4.1)	2.6 (2.2 – 3.3)	2.5 (2.0 – 3.3)	2.0 (1.8 – 2.2)
Observation rate	r	0.024 (0.018 – 0.031)	0.058 (0.047 – 0.071)	0.060 (0.048 – 0.072)	0.029 (0.009 – 0.10)
Fraction of population involved	ρ	74% (69-79)	50% (46-54)	50% (46-54)	34% (8-91)
Initial number of infected individuals	Hi_0	2 (1-8)	5 (0-21)	304 (16-1145)	37 (1-386)
Infectious period in human (days)	γ^{-1}	5.2 (4.0 – 6.5)	5.2 (4.0 – 6.6)	5.2 (4.0 – 6.5)	5.4 (4.1 – 6.7)
Extrinsic incubation period in mosquito (days)	τ^{-1}	10.5 (8.6 – 12.4)	10.5 (8.6 – 12.4)	10.4 (8.5 – 12.5)	10.7 (8.9 – 12.5)
Mosquito lifespan (days)	μ^{-1}	15.4 (11.8 – 19.3)	15.3 (11.6 – 19.2)	14.9 (10.9 – 18.8)	15.3 (11.5 – 19.4)

Table 1 – **Parameter estimations for the Pandey model.** Posterior median (95% credible intervals). All the posterior parameter distributions are presented in Figures 6-9.

LANERI MODEL		Yap	Moorea	Tahiti	New Caledonia
Population size	H	6,892	16,200	178,100	268,767
Basic reproduction number	R_0	2.2 (1.9 – 2.6)	1.8 (1.6 – 2.0)	1.6 (1.5 – 1.7)	1.6 (1.5 – 1.7)
Observation rate	r	0.02 (0.018 – 0.032)	0.057 (0.047 – 0.071)	0.057 (0.048 – 0.068)	0.015 (0.009 – 0.033)
Fraction of population involved	ρ	73% (69 – 78)	51% (46 – 54)	54% (49 – 59)	70% (31 – 100)
Initial number of infected individuals	Hi_0	2 (1 – 10)	9 (1 – 28)	667 (22 – 1570)	82 (2 – 336)
Infectious period in human (days)	γ^{-1}	5.3 (4.0 – 6.6)	5.2 (4.0 – 6.6)	5.1 (4.0 – 6.5)	5.4 (4.0 – 6.7)
Extrinsic incubation period in mosquito (days)	τ^{-1}	10.6 (8.8 – 12.7)	10.6 (8.6 – 12.6)	10.5 (8.5 – 12.4)	10.8 (8.9 – 12.7)

Table 2 – **Parameter estimations for the Laneri model.** Posterior median (95% credible intervals). All the posterior parameter distributions are presented in Figures 10-13.

For the duration of infectious and intrinsic incubation (in human) and extrinsic incubation (in mosquito) periods, the posterior density ressembles the informative prior (cf. Figures 6-13), indicating the models' incapacity to identify properly these parameters without more informative data. Moreover, these parameters have a clear sensitivity (see Materials and methods) and precise field measures are therefore crucial for reliable model predictions.

The fraction ρ of the population involved in the epidemic and the observation rate r display very large credible intervals when seroprevalence is unknown (New Caledonia). They are highly correlated with one another (cf. Figures 17 and 21) and therefore unidentifiable without precise information on seroprevalence.

Discussion

The reproduction number R_0 is a key parameter in epidemiology that characterizes the epidemic dynamics and the initial spread of the pathogen at the start of an outbreak in a susceptible population. R_0 can be used to inform public health authorities on the level of risk posed by an infectious disease, vaccination strategy, and the potential effects of control interventions[19]. In the light of the potential public health crisis generated by the international propagation of ZIKV, characterization of the potential transmissibility of this

pathogen is crucial for predicting epidemic size, rate of spread and efficacy of intervention.

Using data from both surveillance systems and seroprevalence surveys in four different geographical settings across the Pacific,[2;12–15] we have estimated the basic reproductive number R_0 (see Figs 2-3 and Tables 1-2). Our estimate of R_0 obtained by inference based on Particle MCMC[11] has values in the range 1.6 (1.5-1.7) - 3.2 (2.4-4.1). Our R_0 estimates vary notably across settings. Lower and finer R_0 values are found in larger islands. This phenomenon can at least in part be explained by large spatial heterogeneities and higher demographic stochasticity for islands with smaller populations, as well as the influence of stochasticity on biological and epidemiological processes linked to virus transmission. This phenomenon can also be specific to the selection of the studied island or can reflect a highly clustered geographical pattern, the global incidence curve being the smoothed overview of a collection of more explosive small size outbreaks. However, it is notable that the two French Polynesian islands yield similar estimates of R_0 despite differing population sizes. Indeed, other important factors differ among French Polynesia, New Caledonia and Yap, such as the human genetic background and their immunological history linked to the circulation of others arboviruses. Moreover, whilst both New Caledonia and French Polynesia populations were infected by the same ZIKV lineage and transmitted by the same principle vector species, *Aedes aegypti*, the epidemic in Yap occurred much earlier with a different ZIKV lineage[20] and vectored by a different mosquito species *Aedes hensilli*.[21] In French Polynesia, the vector *Aedes polynesiensis* is also present and dominates in Moorea with higher densities than in Tahiti. Finally, different vector control measures have been conducted in the three countries.

To date, studies investigating Zika outbreaks in the Pacific have always estimated R_0 using a deterministic framework. Using a similar version of the Pandey model in French Polynesia, Kucharski et al.[22] estimated R_0 between 1.6 and 2.3 (after scaling to square root for

comparison) for Tahiti and between 1.8 and 2.9 in Moorea. These estimates are slightly lower and less variable than ours. This difference can be explained firstly by the chosen priors on mosquito parameters and secondly because our model includes demographic stochasticity. Moreover, they predicted a seroprevalence rate at the end of the epidemic of 95-97%, far from the 49% measured. In Yap island, a study[23] used a very detailed deterministic mosquito model, and estimated an R_0 for Zika between 2.9 and 8. In this case, our lower and less variable estimates may come from the fact that our model is more parsimonious in the number of uncertain parameters, especially concerning the mosquito population. Finally, a third study[24] relied on another method for R_0 calculation (based only on the early exponential growth rate of the epidemic) in French Polynesia as a whole and in Yap. Again, the obtained parameters are lower than ours in French Polynesia and higher in Yap. In all these studies a deterministic framework is used excluding the possibility of accounting for the high variability of biological and epidemiological processes exacerbated by the small size of the population. In these three studies, like in ours, it is worth noting that little insight is obtained regarding mosquito parameters. Posterior distribution mimics the chosen prior (cf. Figures 6-13). Both the simulation of the epidemics and the estimated R_0 are highly sensitive to the choice of priors on mosquito parameters, for which precise field measures are rare.

In the absence of sufficient data, the modeling of mosquito-borne pathogen transmission is a difficult task due to non-linearity and non-stationarity of the involved processes.[25] This work has then several limitations. First, our study is limited by the completeness and quality of the data, with regard to both incidence and seroprevalence, but, above all, by the scarcity of information available on mosquitoes. Incidence data is aggregated at the island scale and cannot disentangle the effects of geography and observation noise to explain bimodal curves observed in Yap and New Caledonia. Moreover, although all data came from national surveillance systems, we had very little information about the potential degree

of under-reporting. Seroprevalence data were gathered from small sample sizes and were missing in New Caledonia, which leads to strong correlation between the observation rate and the fraction of the population involved in the epidemic. Because of the high proportion of asymptomatic or mildly symptomatic cases, the magnitude of the outbreaks is difficult to evaluate without precise seroprevalence data[26] or detection of mild, asymptomatic or pre-symptomatic infections.[27] Considering vectors, no demographic data were available and this partly explains the large variability of our R_0 estimations.

Secondly, incidence and seroprevalence data were difficult to reconcile; the use of incidence data led to higher infection rates than those observed in the seroprevalence data. This difficulty has been overcome by considering that only a fraction of the population (ρ) is involved in the epidemic and then our model manages to reproduce the observed seroprevalence rate. This exposed fraction could be the result of spatial heterogeneity and high clustering of cases and transmission, as observed for dengue. Finer scale incidence and seroprevalence data would be useful to explore this. Another explanation for higher predicted than observed infection rates could be due to interaction with other flaviviruses. The Zika outbreak was concomitant with dengue outbreaks in French Polynesia[12;13] and New Caledonia.[15] Examples of coinfection have been reported[4] but competition between these close pathogens may also have occurred. Finally, mathematical models with vectorial transmission may tend to estimate high attack rates, sometimes leading to a contradiction between observed incidence and observed seroprevalence. Assumptions on the proportionality between infected mosquitoes and the force of infection, as well as the density-dependence assumption in these models could be questioned. Indeed even if these assumptions are at the heart of the mathematical models of mosquito-borne pathogen transmission[28;29] a recent review,[30] and recent experimental results[31;32] question these important points.

On a final note, the estimates of R_0 for ZIKV are similar to but generally on the lower side of estimates made for two other flaviviruses of medical importance, dengue and Yellow Fever

viruses[33–35], even though caution is needed in the comparison of studies with differing models, methods and data sources. Interventions strategies developed for dengue should thus enable as, if not more effective control of ZIKV, with the caveat that ZIKV remains principally a mosquito-borne pathogen with little epidemiological significance of the sexual transmission route.

In conclusion, using state-of-the art stochastic modeling methods, we have been able to determine estimates of R_0 for ZIKV with an unexpected relationship with population size. Further data from the current Zika epidemic in South America that is caused by the same lineage as French Polynesia will enable us to confirm this relationship. Our study highlights the importance of gathering seroprevalence data, especially for a virus that often leads to an asymptomatic outcome and it would provide a key component for precise quantitative analysis of pathogen propagation to enable improved planning and implemention of prevention and control strategies.

Materials and methods

Data

During the 2007 outbreak that struck Yap, 108 suspected or confirmed Zika cases (16 per 1,000 inhabitants) were reported by reviewing medical records and conducting prospective surveillance between April 1st and July 29th 2007.[2] In French Polynesia, sentinel surveillance recorded more than 8,700 suspected cases (32 per 1,000 inhabitants) across the whole territory between October 2013 and April 2014.[12;13] In New Caledonia, the first Zika case was imported from French Polynesia on 2013 November 12th. Approximately 2,500 cases (9 per 1,000 inhabitants) were reported through surveillance between January (first autochtonous case) and August 2014.[15]

For Yap and French Polynesia, the post-epidemic seroprevalence was assessed. In Yap, a household survey was conducted after the epidemic, yielding an infection rate in the island of 73%.[2] In French Polynesia, three seroprevalence studies were conducted. The first one took place before the Zika outbreak, and concluded that most of the population was naive for Zika virus.[36] The second seroprevalence survey was conducted between February and March 2014, at the end of the outbreak, and reported a seroprevalence rate around 49%.[14] The third one concerned only schoolchildren in Tahiti and was therefore not included in the present study.

Demographic data on population size were based on censuses from Yap[2], French Polynesia[37], and New Caledonia.[38]

Models and inference

Model equations

Although the models are simulated in a stochastic framework, we present them with ordinary differential equations for clarity. The reactions involved in the stochastic models are the same as those governed by the deterministic equations, but the simulation process differs through the use of discrete compartments. It is described in the next section.

The equations describing Pandey model are:

$$
\begin{aligned}
\frac{dH_S}{dt} &= -\beta_H v_I H_S \\
\frac{dH_E}{dt} &= \beta_H v_I H_S - \sigma H_E \\
\frac{dH_I}{dt} &= \sigma H_E - \gamma H_I \\
\frac{dH_R}{dt} &= \gamma H_I \\
\frac{dv_S}{dt} &= \mu - \frac{\beta_V H_I}{N} v_S - \mu v_S \\
\frac{dv_E}{dt} &= \frac{\beta_V H_I}{N} v_S - \tau v_E - \mu v_E \\
\frac{dv_I}{dt} &= \tau v_E - \mu v_I
\end{aligned}
$$

where $v_s = \frac{V_S}{V}$ is the proportion of susceptible mosquitoes, $v_E = \frac{V_E}{V}$ the proportion of exposed mosquitoes, and $v_I = \frac{V_I}{V}$ the proportion of infected mosquitoes. Since we are using a discrete model, we cannot use directly the proportions v_S, v_E and v_I whose values are smaller than one. Therefore, we rescale using $V = H$, which leads to $V'_S = v_S \cdot H$, $V'_E = v_E \cdot H$, and $V'_I = v_I \cdot H$.

In this model, the force of infection for humans is $\lambda_H = \beta_H v_I$, and the force of infection for mosquitoes is $\lambda_V = \beta_V \frac{H_I}{N}$

The equations describing Laneri model are:

$$
\begin{aligned}
\frac{dH_S}{dt} &= -\lambda H_S \\
\frac{dH_E}{dt} &= \lambda H_S - \sigma H_E \\
\frac{dH_I}{dt} &= \sigma H_E - \gamma H_I \\
\frac{dH_R}{dt} &= \gamma H_I \\
\frac{d\kappa}{dt} &= \frac{2\beta H_I \tau}{N} - 2\tau\kappa \\
\frac{d\lambda}{dt} &= 2\tau\kappa - 2\tau\lambda
\end{aligned}
$$

In this model, the role of mosquitoes in transmission is represented only through the delay they introduce during the extrinsic incubation period (EIP, incubation period in the mosquito). For modeling reasons, this delay is included by representing the force of infection from infected humans to susceptible humans with two compartments κ and λ: in this formalism, the duration between the moment when an exposed individual becomes infectious and the moment when another susceptible individual acquires the infection has a gamma distribution of mean τ^{-1}.[17;39;40] Therefore, λ represents the current force of infection for humans $\lambda_H = \lambda$. The compartment κ represents the same force of infection but at a previous stage, reflecting the exposed phase for mosquitoes during the extrinsic incubation period. As an analogy to Pandey model, the force of infection for mosquitoes is $\lambda_V = \frac{\beta H_I \tau}{v_s N}$, and therefore, the parameter β can be interpreted as the product of a transmission parameter β' by the proportion of susceptible mosquitoes: $\beta = v_s \beta'$. The force of infection for mosquitoes is then similar to Pandey's : $\lambda_V = \beta' \tau \frac{H_I}{N}$.

Again, since we are using a discrete model, we cannot use directly the proportions λ and κ whose values are smaller than one. Therefore, we rescale up to a factor N, which leads to $L = \lambda N$ and $K = \kappa N$.

Stochastic framework

Both models are simulated in a stochastic and discrete framework, the Poisson with stochastic rates formulation,[18] to include the uncertainties related to small population size. In this framework, the number of reactions occurring in a time interval dt is approximated by a multinomial distribution. In a model with m possible reactions and c compartments, z_t being the state of the system at time t and θ the model parameters, the probability that each reaction r^k occurs n_k times in dt is calcutated as follows:[41]

$$p(n_1, ...n_m|z_t, \theta) = \prod_{i=1}^{c} \left\{ M_i (1 - \sum_{X(k)=i} p_k)^{\bar{n_i}} \prod_{X(k)=i} (p_k)^{n_k} \right\} + o(dt)$$

with, $z_t^{(i)}$ being the number of individual in compartment i at time t,

- $p_k = \left(1 - exp - \sum_{X(l)=i} r^l(z_t, \theta) z_t^{X(l)} dt\right) \frac{r^k(z_t,\theta)}{\sum_{X(l)=i} r^l(z_t,\theta)}$

- $\bar{n_i} = z_t^{(i)} - \sum_{X(k)=i} n_k$ (the number of individuals staying in compartment i in dt)

- $M_i = \binom{z_t^{(i)}}{n_k \, X(k)=i \, \bar{n_i}}$ (multinomial coefficient)

Observation models

The only observed compartments are the infected humans (incidence measured every week) and the recovered humans (seroprevalence at the end of the outbreak when data is available). In order to link the model to the data, two observation models, for both incidence and seroprevalence data, are needed.

Observation model on incidence data

The observed weekly incidence is assumed to follow a negative binomial distribution[18] whose mean equals the number of new cases predicted by the model times an estimated observation rate r.

The observation rate r accounts for non observed cases, due to non reporting from medical centers, mild symptoms unseen by health system, and asymptomatic infections. Without additional data, it is not possible to make a distinction between these three categories of cases. We also implicitely make the assumption that these cases transmit the disease as much as reported symptomatic cases.

The observation model for incidence data is therefore :

$$Inc_{obs} = NegBin(\phi^{-1}, \frac{1}{1 + \phi r Inc})$$

Inc_{obs} being the observed incidence, and Inc the incidence predicted by the model. The dispersion parameter[18] ϕ is fixed at 0.1.

Observation model on seroprevalence data

Seroprevalence data is fitted for Tahiti, Moorea, and Yap settings. It is assumed that the observed seroprevalence at the end of the epidemic follows a normal distribution with fixed standard deviation, whose mean equals the number of individuals in the H_R compartment predicted by the model.

The observation model for seroprevalence data is therefore :

$$H_R^{obs} = Normal(H_R, \Lambda)$$

329 at the last time step, with notations detailed for each model in Table 3.

Island	Date	Standard deviation Λ	Observed seroprevalence H_R^{obs}
Yap	2007-07-29	150	5005 [2]
Moorea	2014-03-28	325	$0.49 \times 16200 = 7938$ [14]
Tahiti	2014-03-28	3562	$0.49 \times 178100 = 87269$ [14]

Table 3 – **Details of the observation models for seroprevalence**

330 ## Prior distributions

331 Informative prior distributions are assumed for the mosquito lifespan, the duration of
332 infectious period, and both intrinsic and extrinsic incubation periods. The initial numbers of
333 infected mosquitoes and humans are estimated, and the initial number of exposed individuals
334 is set to the initial number of infected to reduce parameter space. We assume that involved
335 populations are naive to Zika virus prior to the epidemic and set the initial number of
336 recovered humans to zero. The other priors and associated references are listed in Table 4.

337

Parameters		**Pandey model**	**Laneri model**	**References**
R_0^2	squared basic reproduction number	Uniform[0, 20]	Uniform[0, 20]	assumed
β_V	transmission from human to mosquito	Uniform[0, 10]	.	assumed
γ^{-1}	infectious period (days)	Normal(5, 1) in [4, 7]	Normal(5, 1) in [4, 7]	13
σ^{-1}	intrinsic incubation period (days)	Normal(4, 1) in [2, 7]	Normal(4, 1) in [2, 7]	42–44
τ^{-1}	extrinsic incubation period (days)	Normal(10.5, 1) in [4, 20]	Normal(10.5, 1) in [4, 20]	45;46
μ^{-1}	mosquito lifespan (days)	Normal(15, 2) in [4, 30]	.	47;48
ρ	fraction of population involved	Uniform[0, 1]	Uniform[0, 1]	
Initial conditions (t=0)		**Pandey model**	**Laneri model**	
$H_{I\,0}$	Infected humans	Uniform[10^{-6}, 1]·N	Uniform[10^{-6}, 1]·N	
$H_{E\,0}$	exposed humans	$H_{I\,0}$	$H_{I\,0}$	
$H_{R\,0}$	Recovered humans	0	0	
	Infected vectors	$V_{I\,0}$=Uniform[10^{-6}, 1]·H	L_0=Uniform[10^{-6}, 1]·N	
	exposed vectors	$V_{E\,0} = V_{I\,0}$	K_0=L_0	

Local conditions		**Yap**	**Moorea**	**Tahiti**	**New Caledonia**	**References**
r	Observation rate	Uniform[0, 1]	Uniform[0, 1]	Uniform[0, 0.3]	Uniform[0, 0.23]	13;15;49
H	Population size	6,892	16,200	178,100	268,767	2;37;38

Table 4 – **Prior distributions of parameters**. "Uniform[0,20]" indicates a uniform distribution between in the range [0,20]. "Normal(5,1) in [4,7]" indicates a normal distribution with mean 5 and standard deviation 1, restricted to the range [4,7].

The range for the prior on observation rate is reduced for Tahiti and New Caledonia, in order to reduce the parameter space and facilitate convergence. In both cases, we use the information provided with the data source. In French Polynesia, 8,750 cases we reported, but according to local health authorities, more than 32,000 people would have attended health facilities for Zika[13] ($8750/32000 \leq 0.3$). In New Caledonia, approximately 2,500 cases were reported but more than 11,000 cases were estimated by heath authorities[49] ($2500/11000 \leq 0.23$). In both cases, these extrapolations are lower bounds on the real number of cases (in particular, they do not estimate the number of asymptomatic infections), and therefore can be used as upper bounds on the observation rate.

Estimations

Inference with PMCMC

The complete model is represented using the state space framework, with two equation systems: the transition equations refer to the transmission models, and the measurement equations are given by the observation models.

In a deterministic framework, this model could be directly estimated using MCMC, with a Metropolis-Hastings algorithm targeting the posterior distribution of the parameters. This algorithm would require the calculation of the model likelihood at each iteration.

In our stochastic framework, the model output is given only through simulations and the likelihood is intractable. In consequence, estimations are performed with the PMCMC algorithm (particle Markov Chain Monte Carlo [11]), in the PMMH version (particle marginal Metropolis-Hastings). This algorithm uses the Metropolis-Hastings structure, but replaces the real likelihood by its estimation with Sequential Monte Carlo (SMC).

Algorithm 1 PMCMC[11] (PMMH version, as in SSM[41])

In a model with n observations and J particles.

$q(.|\theta^{(i)})$ is the transition kernel.

1: Initialize $\theta^{(0)}$.

2: Using SMC algorithm, compute $\hat{p}(y_{1:n}|\theta^{(0)})$ and sample $x^*_{0:n}$ from $\hat{p}(x_{0:n}|y_{1:n}, \theta^{(0)})$.

3: **for** $i = 1...N^\theta$ **do**

4: Sample θ^* from $q(.|\theta^{(i)})$

5: Using SMC algorithm, compute $L(\theta^*) = \hat{p}(y_{1:n}|\theta^*)$ and sample $x^*_{0:n}$ from $\hat{p}(x_{0:n}|y_{1:n}, \theta^*)$

6: Accept θ^* (et $x^*_{0:n}$) with probability $1 \wedge \frac{L(\theta^{(i)})q(|\theta^*)}{L(\theta^*)q(\theta^*|\theta^{(i)})}$

7: If accepted, $\theta^{(i+1)} = \theta^*$ and $x^{(i+1)}_{0:n} = x^*_{0:n}$. Otherwise $\theta^{(i+1)} = \theta^{(i)}$ and $x^{(i+1)}_{0:n} = x^{(i)}_{0:n}$.

8: **end for**

SMC[50] is a filtering method that enables to recover the latent variables and estimate the likelihood for a given set of parameters. The data is treated sequentially, by adding one more data point at each iteration. The initial distribution of the state variables is approximated by a sample a particles, and from one iteration to the next, all the particles are projected according to the dynamic given by the model. The particles receive a weight according to the quality of their prediction regarding the observations. Before the next iteration, all the particles are resampled using these weights, in order to eliminate low weight particles and concentrate the computational effort in high probability regions. Model likelihood is also computed sequentially at each iteration[41;51].

Algorithm 2 SMC (*Sequential Monte Carlo*, as implemented in SSM[41])

In a model with n observations and J particles.

L is the model likelihood $p(y_{1:T}|\theta)$. $W_k^{(j)}$ is the weight and $x_k^{(j)}$ is the state associated to particle j at iteration k.

Set $L = 1$, $W_0^{(j)} = 1/J$.
Sample $(x_0^{(j)})_{j=1:J}$ from $p(x|\theta_0)$.
for $k = 0 : n - 1$ **do**
    for $j = 0 : J$ **do**
        Sample $(x_{k+1}^{(j)})_{j=1:J}$ from $p(x_{k+1}|x_k, \theta)$
        Set $\alpha^{(j)} = h(y_{k+1}, x_{k+1}^{(j)}, \theta)$
    end for
    Set $W_{k+1}^{(j)} = \frac{\alpha^{(j)}}{\sum_{l=1}^{J} \alpha^{(l)}}$ and $L = L\frac{1}{J}\sum_j \alpha^{(l)}$
    Resample $(x_{0:k+1}^{(j)})_{j=1:J}$ from $W_{k+1}^{(j)}$
end for

A gaussian kernel $q(.|\theta^{(i)})$ is used in the PMCMC algorithm, with mean $\theta^{(i)}$ and fixed variance Σ^q (random walk Metropolis Hastings).

Initialization

PMCMC algorithm is very sensitive to initialization of both the parameter values $\theta^{(0)}$ and the covariance matrix Σ^q. Several steps of initialization are therefore used.

Firstly, parameter values are initialized by maximum likelihood through simplex algorithm on a deterministic version of the model. We apply the simplex algorithm to a set of 1000 points sampled in the prior distributions and we select the parameter set with the highest likelihood.

Secondly, in order to initialise the covariance matrix, an adaptative MCMC (Metropolis Hastings) framework is used[41;52]. It uses the empirical covariance of the chain $\Sigma^{(i)}$, and aims to calibrate the acceptance rate of the algorithm to an optimal value. The transition kernel is also mixed (with a probability $\alpha = 0.05$) with another gaussian using the identity matrix to improve mixing properties.

$$q^A(.|x^{(i)}) = \alpha N(x^{(i)}, \lambda \frac{2.38^2}{d} Id) + (1-\alpha) N(x^{(i)}, \frac{2.38^2}{d} \Sigma^{(i)})$$

The parameter λ is approximated by successive iterations using the empirical acceptance rate of the chain.

$$\lambda_{i+1} = \lambda_i a^i (AccRate_i - 0.234)$$

The adaptative PMCMC algorithm itself may have poor mixing properties without initialization. A first estimation of the covariance matrix is computed using KMCMC algorithm.[41] In the KMCMC algorithm, the model is simulated with stochastic differential equations (intermediate between deterministic and Poisson with stochastic rates frameworks) and the SMC part of the adaptative PMCMC is replaced by the extended Kalman filter. When convergence is reached with KMCMC, then, adaptative PMCMC is used.

The PMCMC algorithm is finally applied on the output of the adaptative PMCMC, using 50,000 iterations and 10,000 particles. Calculations are performed with SSM software[41] and R version 3.2.2.

R_0 calculation

R_0 is calculated using the Next Generation Matrix approach[9] (NGM).

R_0 calculation in Pandey model

$$F = \begin{pmatrix} 0 & 0 & 0 & \beta_H \\ 0 & 0 & 0 & 0 \\ 0 & \beta_V & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad V = \begin{pmatrix} -\sigma & 0 & 0 & 0 \\ \sigma & -\gamma & 0 & 0 \\ 0 & 0 & -(\mu+\tau) & 0 \\ 0 & 0 & \tau & -\mu \end{pmatrix}$$

Then we have,

$$V^{-1} = \begin{pmatrix} -1/\sigma & 0 & 0 & 0 \\ -1/\gamma & -1/\gamma & 0 & 0 \\ 0 & 0 & -1/(\mu+\tau) & 0 \\ 0 & 0 & -\tau/[\mu(\tau+\mu)] & -1/\mu \end{pmatrix}$$

and

$$FV^{-1} = \begin{pmatrix} 0 & 0 & -\beta_H\tau/[\mu(\tau+\mu)] & -\beta_H/\mu \\ 0 & 0 & 0 & 0 \\ -\beta_V/\gamma & -\beta_V/\gamma & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We calculate the eigen values α of $-FV^{-1}$:

$$\begin{vmatrix} -\alpha & 0 & \beta_H\tau/[\mu(\tau+\mu)] & \beta_H/\mu \\ 0 & -\alpha & 0 & 0 \\ \beta_V/\gamma & \beta_V/\gamma & -\alpha & 0 \\ 0 & 0 & 0 & -\alpha \end{vmatrix} = \alpha^2\left(\alpha^2 - \frac{\beta_H\beta_V\tau}{\gamma\mu(\tau+\mu)}\right) = 0$$

Then $\alpha = 0$ or $\alpha = \pm\sqrt{\frac{\beta_H\beta_V\tau}{\gamma\mu(\tau+\mu)}}$ and the highest eigen value is $R_0 = \sqrt{\frac{\beta_H\beta_V\tau}{\gamma\mu(\tau+\mu)}}$.

This formula defines R_0 as "the number of secondary cases per generation"[53]: i.e R_0 can be written as the geometric mean $R_0 = \sqrt{R_0^v R_0^h}$, where R_0^v is the number of infected mosquitoes after the introduction of one infected human in a naive population, and R_0^h is the number of infected humans after the introduction of one infected mosquito in a naive population. With this definition, herd immunity is reached when $(1 - R_0^{-2})$ of the population is vaccinated[53].

R_0 calculation in Laneri model

Following the analogy with Pandey model, we compute the spectral radius of the NGM for the Laneri model.

$$F = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & \beta\tau & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad V = \begin{pmatrix} -\sigma & 0 & 0 & 0 \\ \sigma & -\gamma & 0 & 0 \\ 0 & 0 & -\tau & 0 \\ 0 & 0 & \tau & -\tau \end{pmatrix}$$

Then we have,

$$V^{-1} = \begin{pmatrix} -1/\sigma & 0 & 0 & 0 \\ -1/\gamma & -1/\gamma & 0 & 0 \\ 0 & 0 & -1/\tau & 0 \\ 0 & 0 & -1/\tau & -1/\tau \end{pmatrix}$$

and

$$FV^{-1} = \begin{pmatrix} 0 & 0 & -1/\tau & -1/\tau \\ 0 & 0 & 0 & 0 \\ -\beta\tau/\gamma & -\beta\tau/\gamma & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We calculate the eigen values α of $-FV^{-1}$:

$$\begin{vmatrix} -\alpha & 0 & 1/\tau & 1/\tau \\ 0 & -\alpha & 0 & 0 \\ \beta\tau/\gamma & \beta\tau/\gamma & -\alpha & 0 \\ 0 & 0 & 0 & -\alpha \end{vmatrix} = \alpha^2 \left(\alpha^2 - \frac{\beta\tau}{\gamma\tau} \right) = 0$$

Then $\alpha = 0$ or $\alpha = \pm\sqrt{\frac{\beta}{\gamma}}$ and the highest eigen value is $\alpha_R = \sqrt{\frac{\beta}{\gamma}}$.

Because λ and κ can be seen as parameters rather than state variables, the interpretation of the spectral radius as the R_0 of the model is not straightforward. Therefore, we computed the R_0 of the model through simulations, by counting the number of secondary infections after the introduction of a single infected individual in a naive population. Since Laneri model is considered here as a vector model, the number of infected humans after the introduction of a single infected is considered as R_0^2. We simulated 1000 deterministic

trajectories, using parameter values sampled in the posterior distributions for all parameters except initial conditions. With this method, the confidence intervals for number of infected humans (R_0^2) are similar to the ones of α_R^2 estimated by the model. As a consequence, R_0 was approximated by the spectral radius of the NGM in our results with our stochastic framework (cf. Table 5).

As a robustness check, the same method was applied to Pandey model : the confidence intervals for the number of secondary cases in simulations are very similar to the ones of R_0^2 (cf. Table 5).

	Pandey model	**Laneri model**
Yap	3.1 (2.5-4.3)	2.2 (1.9-2.6)
Moorea	2.6 (2.2-3.3)	1.8 (1.6-2.0)
Tahiti	2.5 (2.0-3.3)	1.6 (1.5-1.7)
New Caledonia	2.0 (1.8-2.2)	1.6 (1.5-1.7)

Table 5 – **Square root of the number of secondary cases after the introduction of a single infected individual in a naive population.** Median and 95% credible intervals of 1000 deterministic simulations using parameters from the posterior distribution.

Sensitivity analysis

In order to analyse the influence of parameter values on the model's outputs, a sensitivity analysis was performed, using LHS/PRCC technique[54], on Tahiti example. Similar results were obtained for the other settings. Three criteria were retained as outputs for the analysis: the seroprevalence at the last time point, the intensity of the peak of the outbreak and the date of the peak. We used uniform distributions for all parameters, which are listed in Tables 6 and 7. For model parameters, we used the same range as for the prior distribution. For initial conditions, the observation rate r and the fraction involved in the epidemic ρ, we used the 95% confidence interval obtained by PMCMC, in order to avoid unrealistic scenarios.

449

Parameters		Seroprevalence		Peak intensity		Peak date	
	Distribution	PRCC	p-value	PRCC	p-value	PRCC	p-value
Model parameters							
R_0^2	Uniform[0,20]	0.88	<0.001	0.91	<0.001	-0.53	<0.001
β_V	Uniform[0.1,10]	-0.69	<0.001	-0.77	<0.001	0.29	<0.001
γ^{-1}	Uniform[4,7]	-0.25	<0.001	0.09	0.003	0.20	<0.001
σ^{-1}	Uniform[2,7]	-0.03	0.324	-0.12	<0.001	0.18	<0.001
τ^{-1}	Uniform[4,20]	0.00	0.921	-0.05	0.102	0.08	0.012
μ^{-1}	Uniform[4,30]	-0.61	<0.001	-0.75	<0.001	0.51	<0.001
Initial conditions							
Hi_0	Uniform[2.10^{-5},0.011]	-0.03	0.403	-0.04	0.204	0.03	0.417
Vi_0	Uniform[10^{-4},0.034]	0.06	0.044	-0.03	0.290	-0.24	<0.001
Fraction involved and observation model							
ρ	Uniform[0.46,0.54]	0.50	<0.001	0.14	<0.001	-0.03	0.307
r	Uniform[0.048,0.072]	0.02	0.578	0.03	0.383	0.04	0.250

Table 6 – **Sensitivity analysis in Pandey model. Tahiti island.** 1000 parameter sets were sampled with latin hypercube sampling (LHS), using "lhs" R package[55]. On each parameter set, the model was simulated deterministically. PRCC were computed using the "sensitivity" R package.[56] P-values were calculated using the Student distribution approximation provided by Blower and Dowlatabadi.[54]

450 For all criteria, the key parameters in both models are transmission parameters (R_0 and β_V).
451 High values for R_0 are positively correlated with a large seroprevalence and a high and early
452 peak. On the contrary, high values for the parameters introducing a delay in the model, the
453 incubation periods in human (σ^{-1}) and in mosquito (τ^{-1}), are associated with a lower and
454 later peak, and have no significant effect on seroprevalence. Moreover, the simulations are
455 clearly sensitive to the other model parameters, in particular the mosquito lifespan (μ^{-1}) in
456 Pandey model.
457 Concerning other parameters, the initial conditions have a noticeable effect on the date
458 of the peak only. As expected, the fraction involved in the epidemic (ρ) influences the
459 magnitude of the outbreak, by calibrating the proportion of people than can be infected, but
460 it has no significant effect on the timing of the peak.

461

Parameters		Seroprevalence		Peak intensity		Peak date	
	Distribution	PRCC	p-value	PRCC	p-value	PRCC	p-value
Model parameters							
R_0^2	Uniform[0,20]	0.62	<0.001	0.93	<0.001	-0.50	<0.001
γ^{-1}	Uniform[4,7]	0.01	0.731	0.62	<0.001	0.15	<0.001
σ^{-1}	Uniform[2,7]	-0.03	0.373	-0.54	<0.001	0.21	<0.001
τ^{-1}	Uniform[4,20]	-0.03	0.323	-0.70	<0.001	0.47	<0.001
Initial conditions							
Hi_0	Uniform[10^{-5},0.015]	0.05	0.135	0.02	0.622	-0.32	<0.001
L_0	Uniform[2.10^{-5},0.004]	0.05	0.133	0.00	0.930	-0.16	<0.001
Fraction involved and observation model							
ρ	Uniform[0.49,0.59]	0.80	<0.001	0.34	<0.001	0.02	0.558
r	Uniform[0.048,0.068]	-0.01	0.727	0.01	0.738	-0.02	0.635

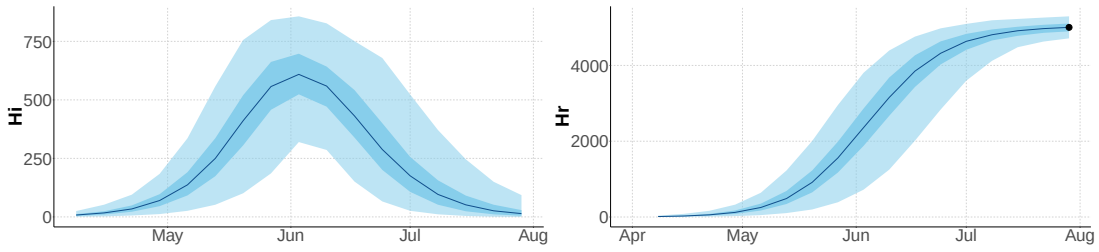
Table 7 – **Sensitivity analysis in Laneri model. Tahiti island.** 1000 parameter sets were sampled with latin hypercube sampling (LHS), using "lhs" R package[55]. On each parameter set, the model was simulated deterministically. PRCC were computed using the "sensitivity" R package.[56] P-values were calculated using the Student distribution approximation provided by Blower and Dowlatabadi.[54]

462 # Complementary results

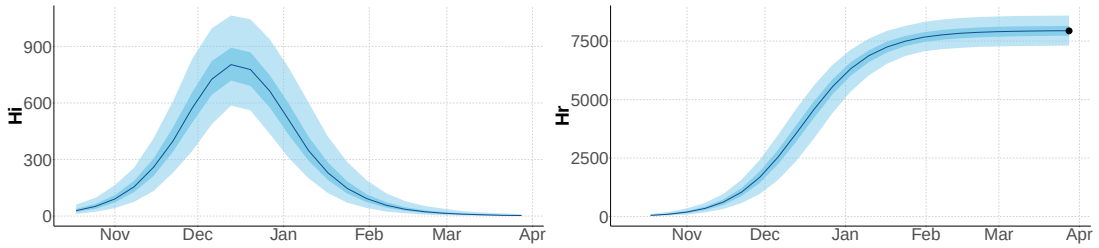
463 These complementary results include PMCMC results for both models in the four settings:
464 the epidemic trajectories regarding the human compartments for infected and recovered
465 individuals (Figures 4-5), the detailed posterior distributions for all parameters (Figures
466 6-13) and correlation plots for all models (Figures 14-21).

467

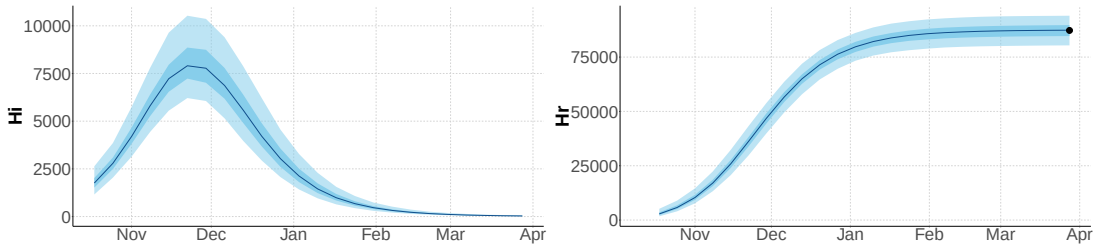
468 ## Infected and recovered



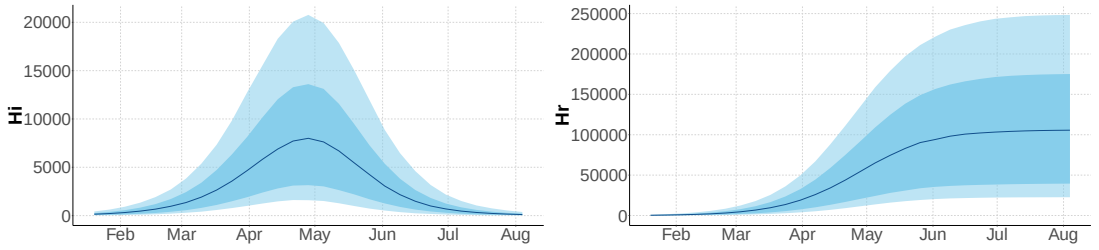
(a)



(b)



(c)



(d)

Figure 4 – **Infected and recovered humans evolution during the outbreak with Pandey model.** Simulations from the posterior density: posterior median (solid line), 95% and 50% credible intervals (shaded blue areas) and observed seroprevalence (black dots). First column: Infected humans (H_I). Second column: Recovered humans (H_R). a) Yap. b) Moorea. c) Tahiti. d) New Caledonia.

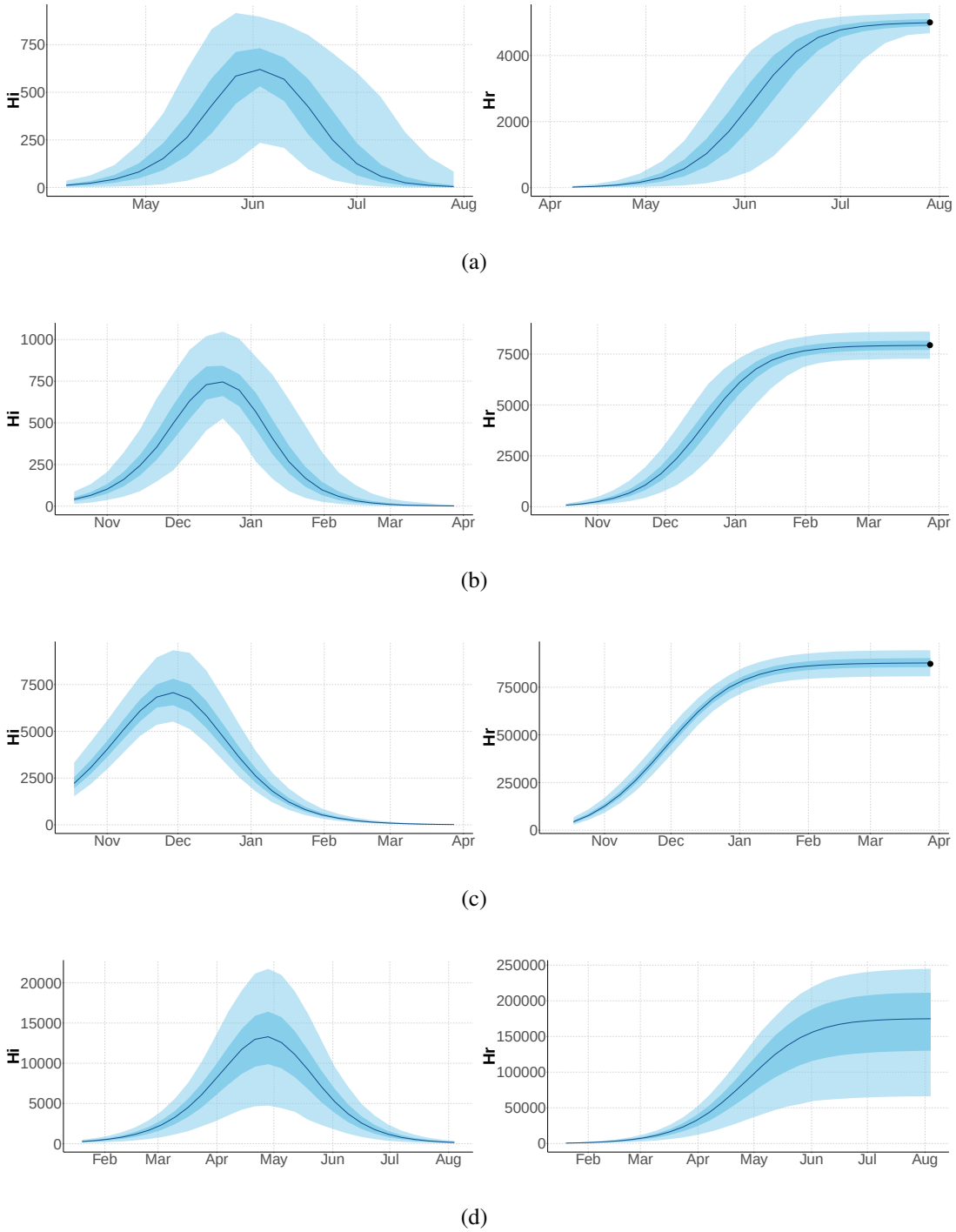


Figure 5 – **Infected and recovered humans evolution during the outbreak with Laneri model.** Simulations from the posterior density: posterior median (solid line), 95% and 50% credible intervals (shaded blue areas) and observed seroprevalence (black dots). First column: Infected humans (H_I). Second column: Recovered humans (H_R). a) Yap. b) Moorea. c) Tahiti. d) New Caledonia.

Posterior distributions

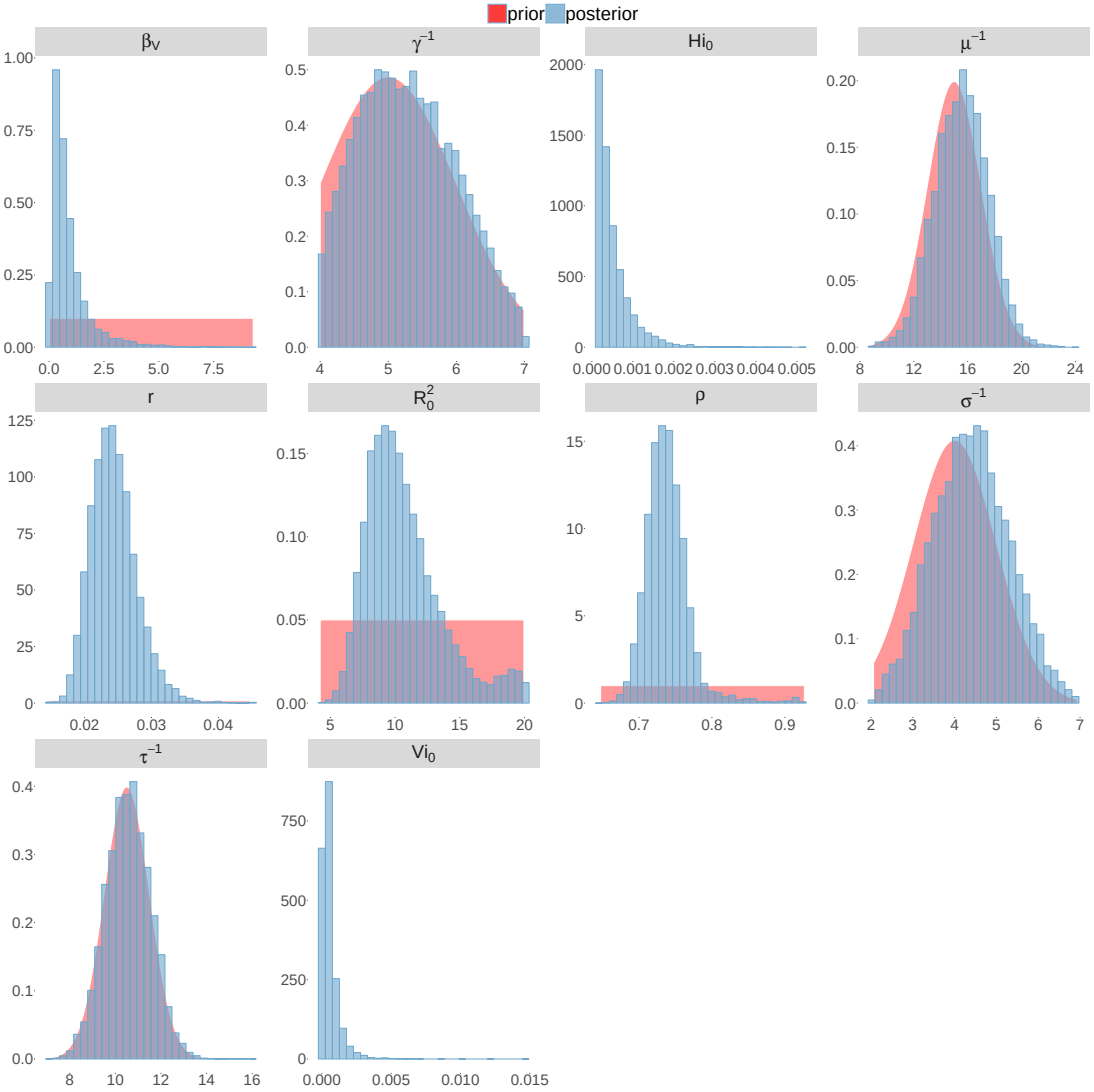


Figure 6 – **Posterior distributions. Pandey model, Yap island.**

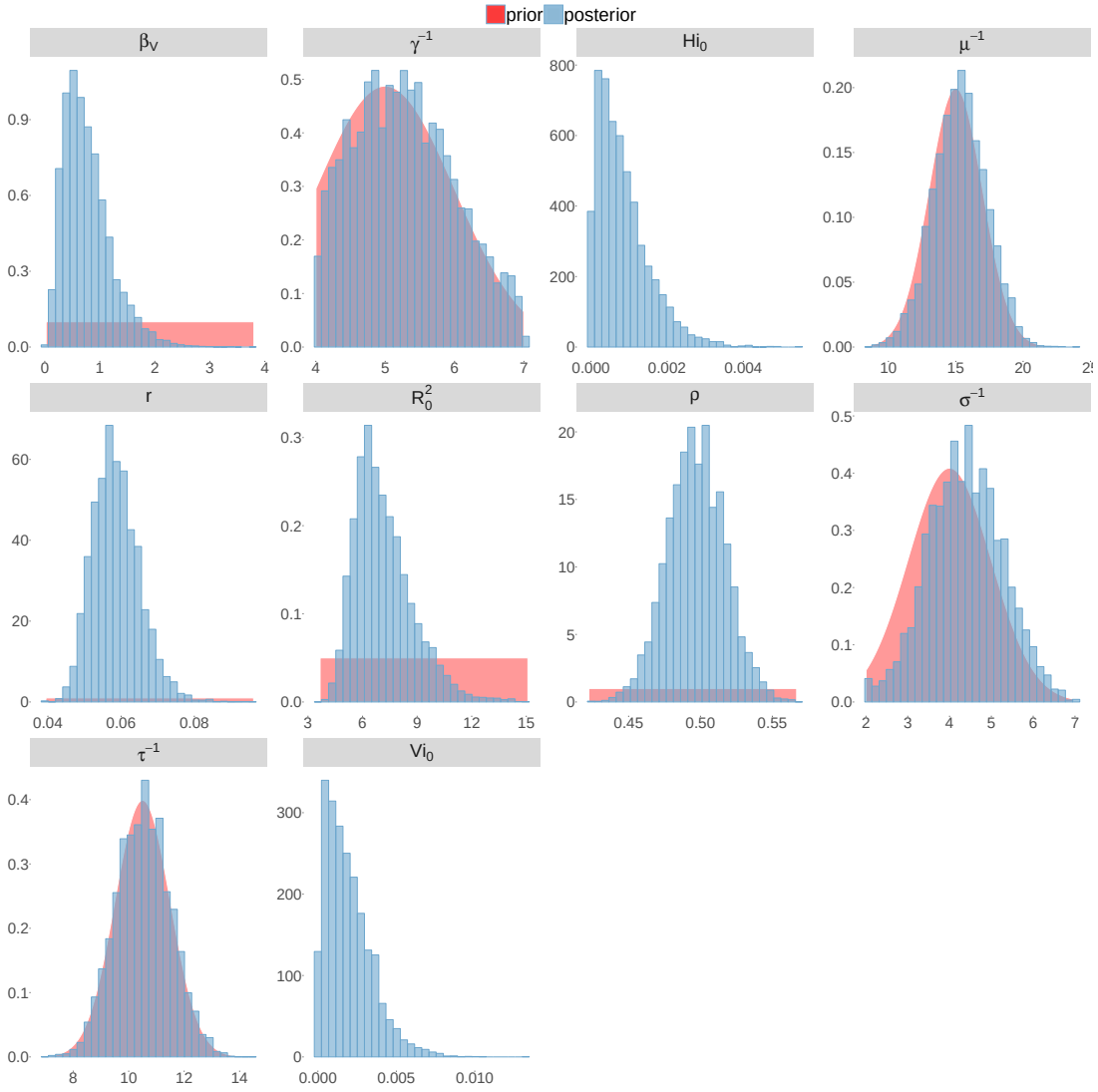


Figure 7 – **Posterior distributions. Pandey model, Moorea island.**

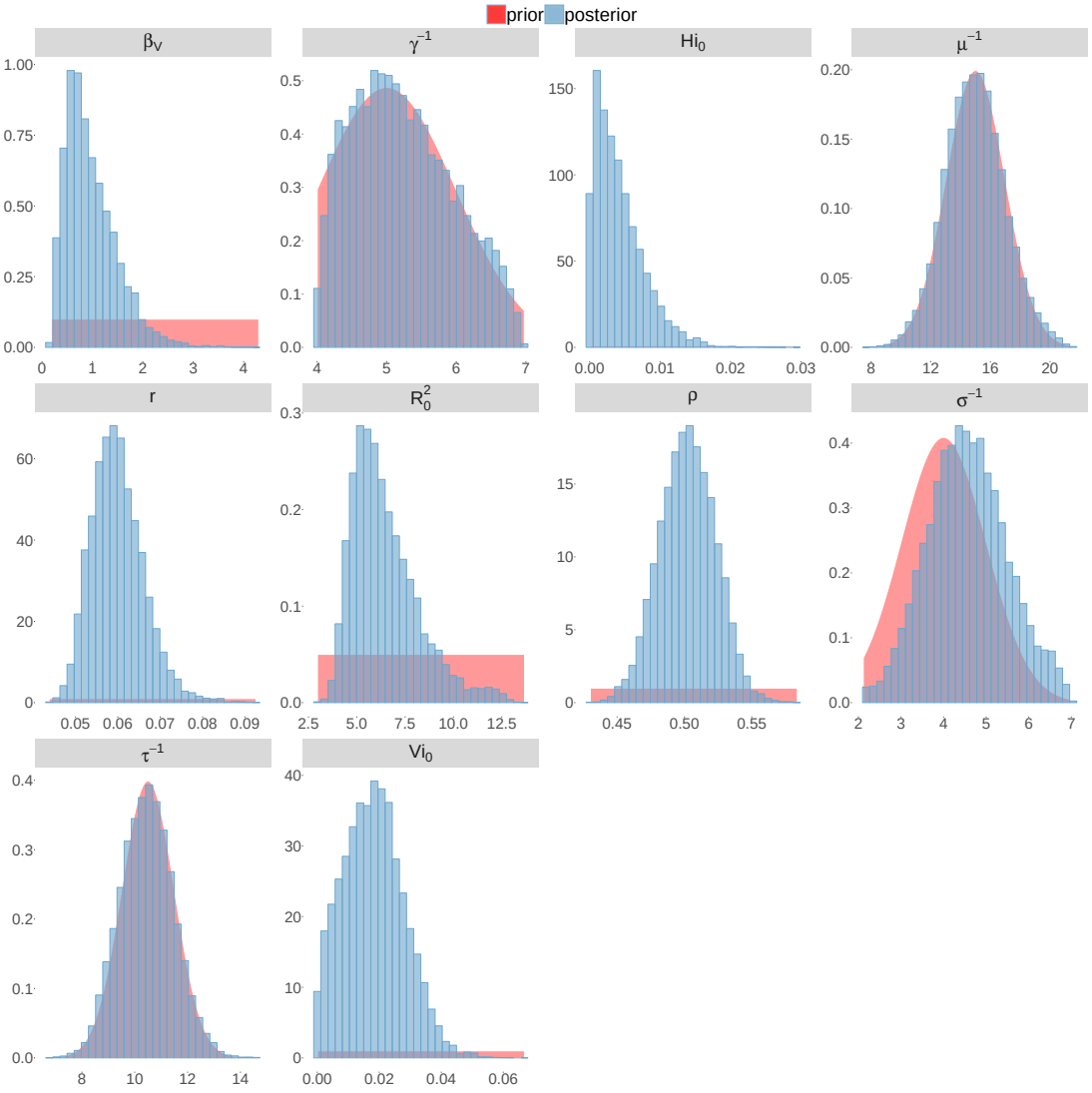


Figure 8 – **Posterior distributions. Pandey model, Tahiti island.**

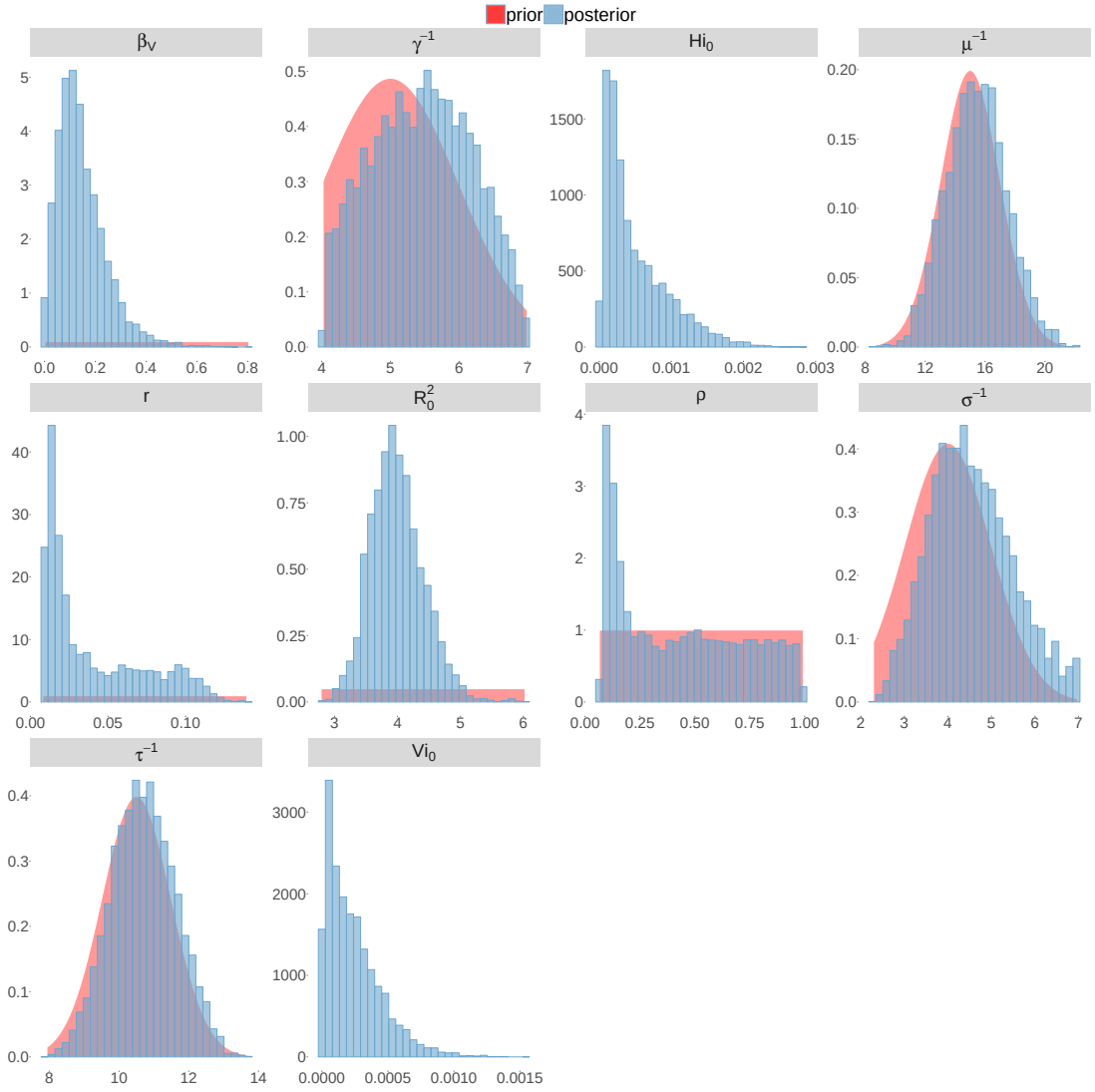


Figure 9 – **Posterior distributions. Pandey model, New Caledonia.**

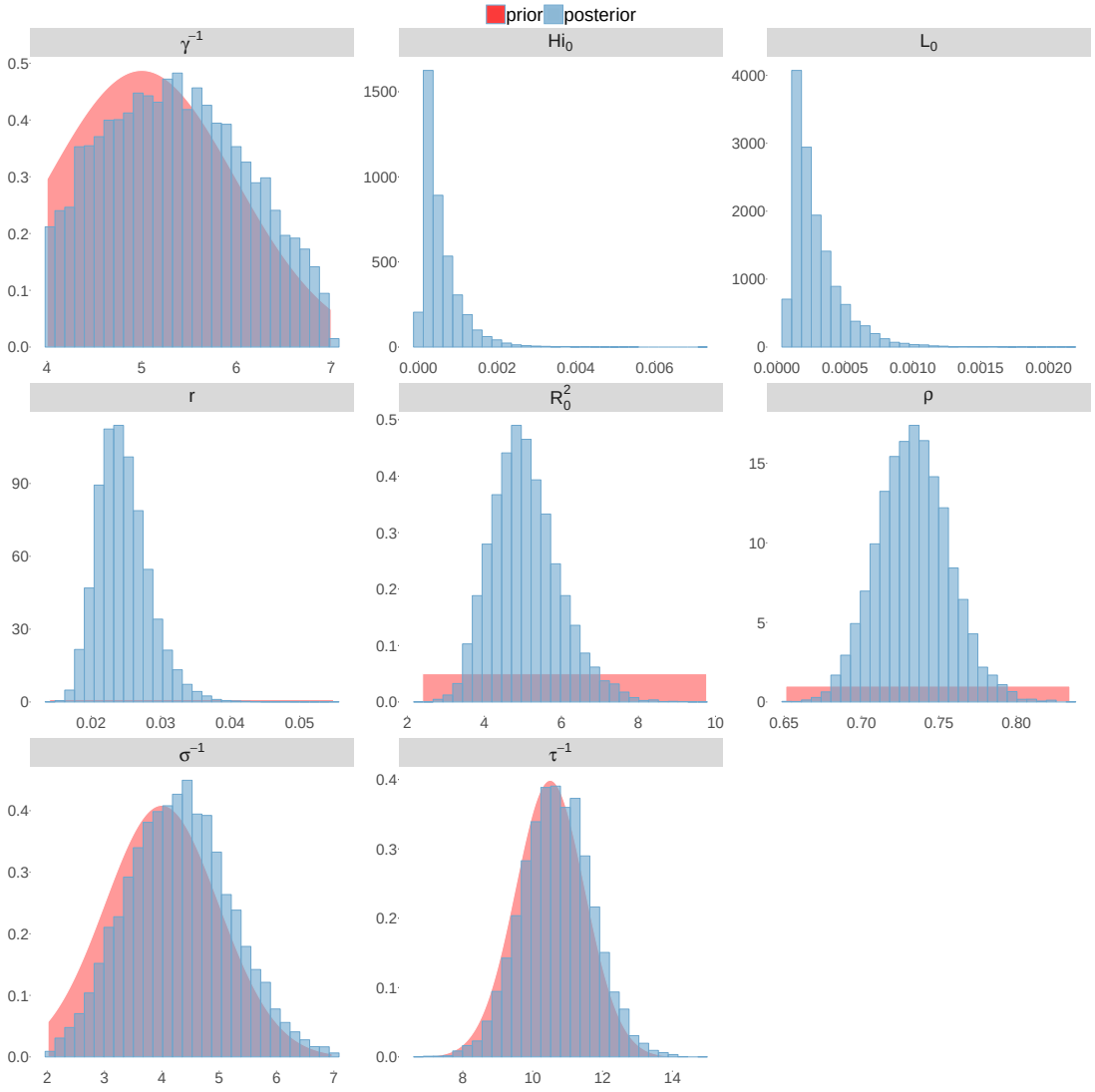


Figure 10 – **Posterior distributions. Laneri model, Yap island.**

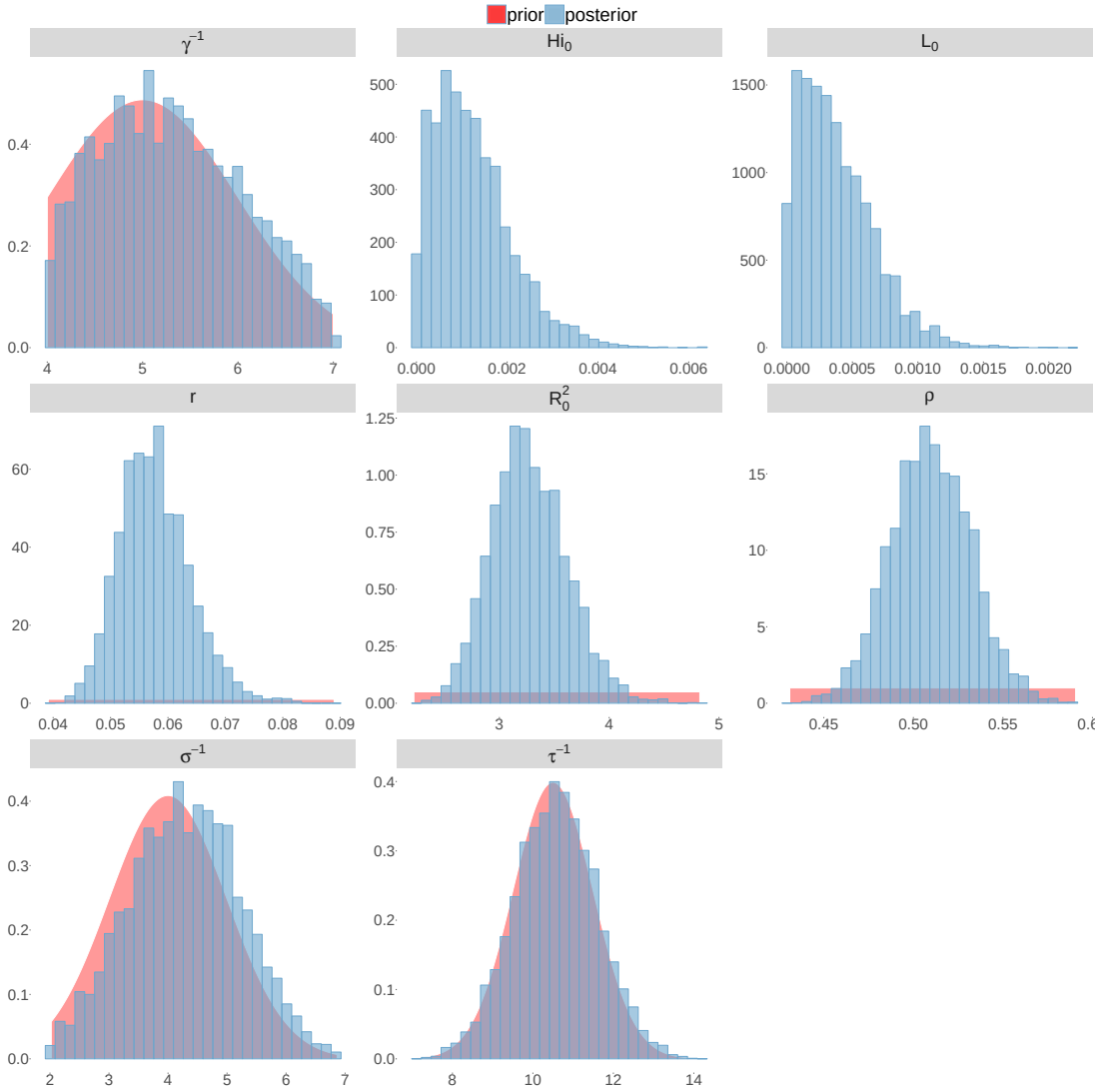


Figure 11 – **Posterior distributions. Laneri model, Moorea island.**

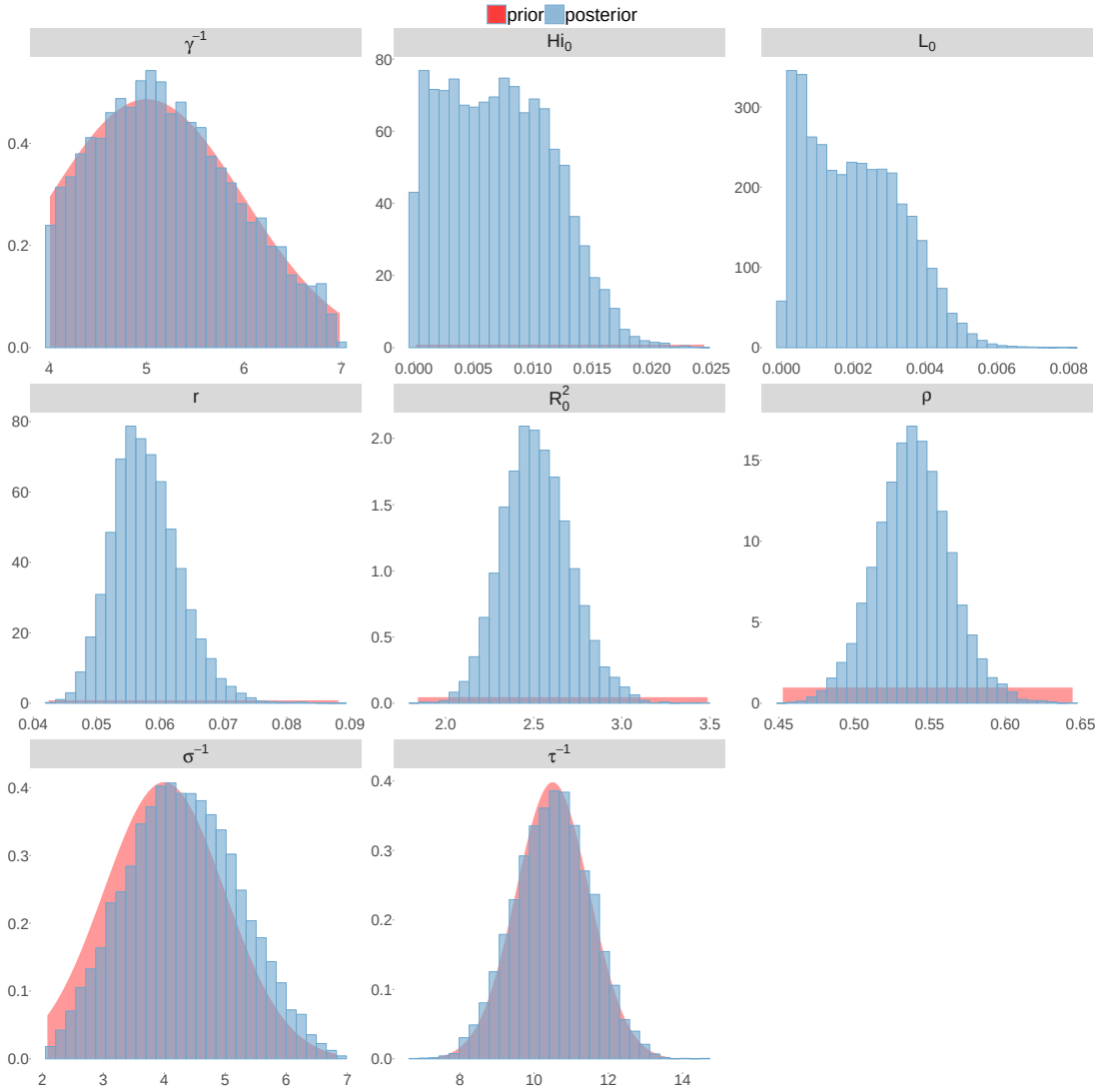


Figure 12 – **Posterior distributions. Laneri model, Tahiti island.**

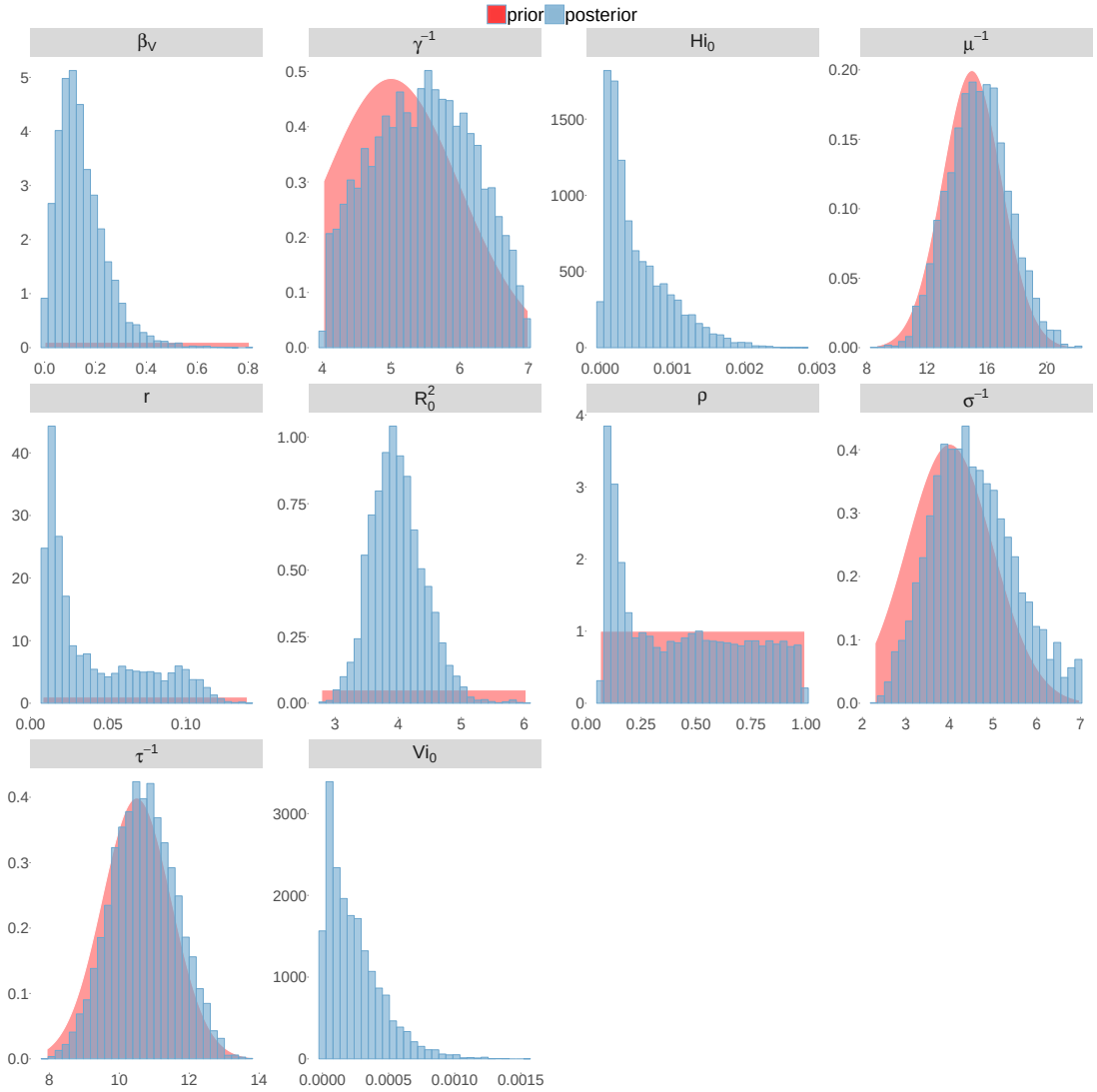


Figure 13 – **Posterior distributions. Laneri model, New Caledonia.**

Correlation between estimated parameters

The marginal posterior densities (Figures 6-13) do not indicate the correlation between parameters, i.e when the observed value of one parameter is highly influenced by the value of another one. In some cases, this phenomenon reveals identifiability issues : the model manages to estimate only a pair of parameters but cannot identify each one separately. In our case, the observation rate and the fraction of the population involved in the epidemic are strongly negatively correlated when no information is provided on seroprevalence (Figures 9 and 13).

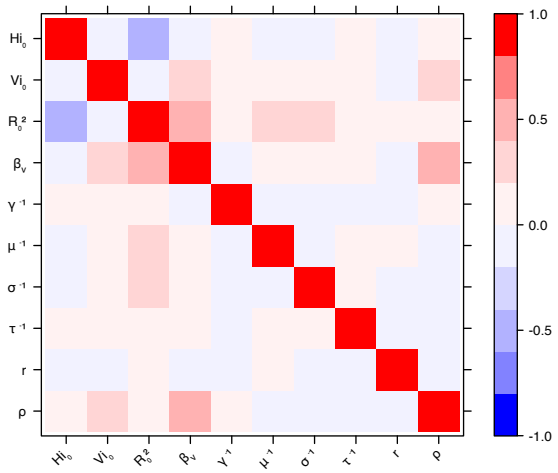


Figure 14 – **Correlation plot of MCMC output. Pandey model, Yap island.**

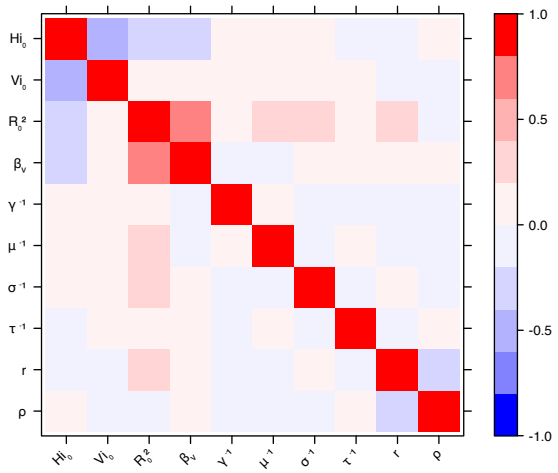


Figure 15 – **Correlation plot of MCMC output. Pandey model, Moorea island.**

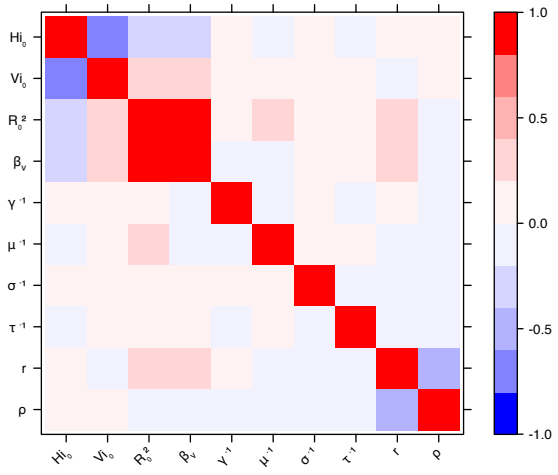


Figure 16 – **Correlation plot of MCMC output. Pandey model, Tahiti island.**

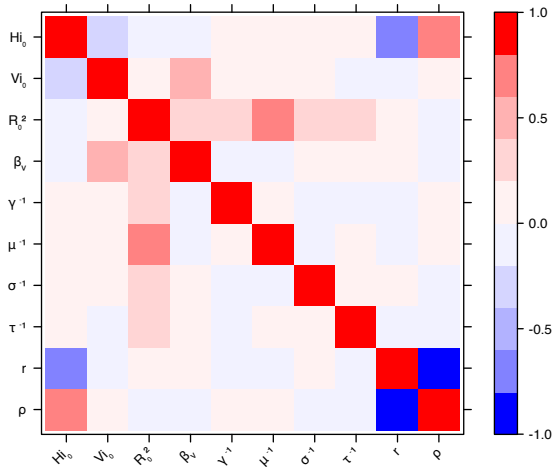


Figure 17 – **Correlation plot of MCMC output. Pandey model, New Caledonia.**

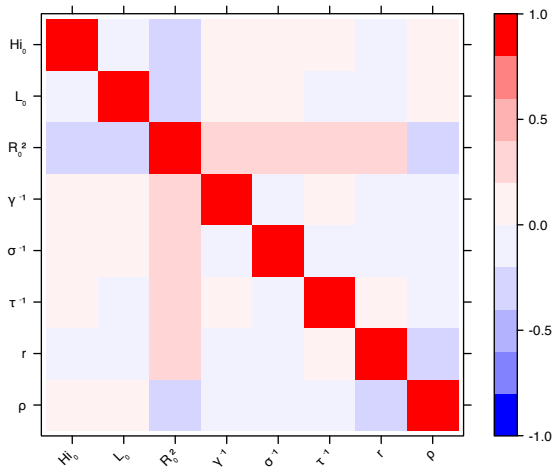


Figure 18 – **Correlation plot of MCMC output. Laneri model, Yap island.**

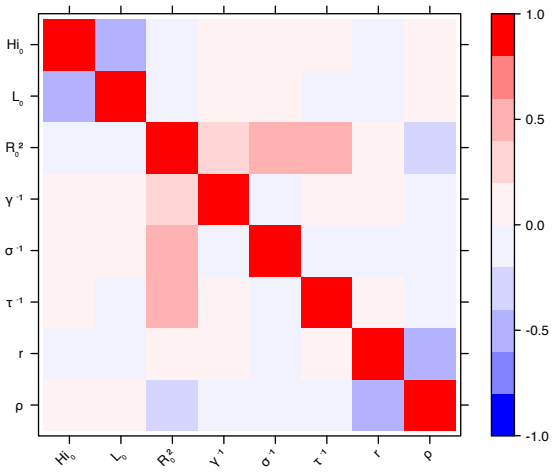


Figure 19 – **Correlation plot of MCMC output. Laneri model, Moorea island.**

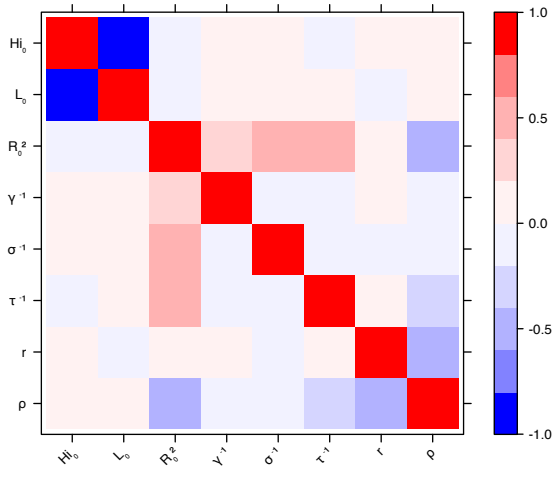


Figure 20 – **Correlation plot of MCMC output. Laneri model, Tahiti island.**

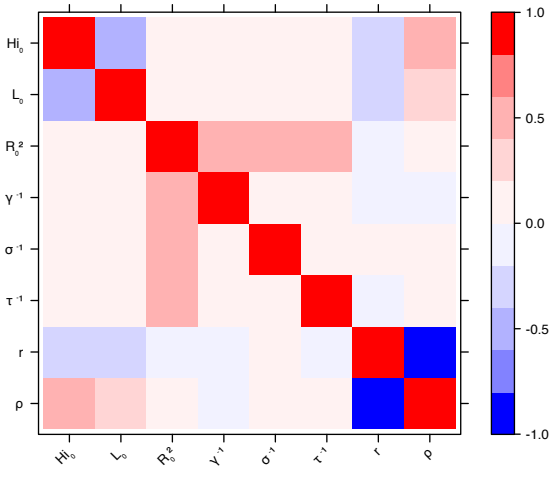


Figure 21 – **Correlation plot of MCMC output. Laneri model, New Caledonia.**

Acknowledgments

CC, DGS and BC are partially supported by the "Pépiniere interdisciplinaire Eco-Evo-Devo" from the Centre National de la Recherche Scientifique (CNRS). The research leading to

these results has also received funding from the European Commission Seventh Framework Program [FP7/ 2007-2013] for the DENFREE project under Grant Agreement n° 282 378. The funders played no role in the study design, data collection, analysis, or preparation of the manuscript.

Contributors

CC and BC designed the study, CC, DGS and BC contributed to the numerical part of the study, and all the authors participated in the interpretation of the results and in the writing of the manuscript.

Declaration of interests

The authors have declared that no competing interests exist.

References

[1] WHO | WHO statement on the first meeting of the International Health Regulations (2005) (IHR 2005) Emergency Committee on Zika virus and observed increase in neurological disorders and neonatal malformations;. Available from: http://www.who.int/mediacentre/news/statements/2016/1st-emergency-committee-zika/en/ [accessed 2016-06-09].

[2] Duffy MR, Chen TH, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on Yap Island, federated states of Micronesia. *New England Journal of Medicine*. 2009;360(24):2536–2543. Available from: http://www.nejm.org/doi/full/10.1056/NEJMoa0805715 [accessed 2016-02-22].

[3] Cao-Lormeau VM, Roche C, Teissier A, Robin E, Berry AL, Mallet HP, et al. Zika

Virus, French Polynesia, South Pacific, 2013. *Emerging Infectious Diseases*. 2014 Jun;20(6):1084–1086. Available from: http://wwwnc.cdc.gov/eid/article/20/6/14-0138_article.htm [accessed 2016-05-23]. doi:10.3201/eid2006.140138.

[4] Dupont-Rouzeyrol M, O'Connor O, Calvez E, Daurès M, John M, Grangeon JP, et al. Co-infection with Zika and Dengue Viruses in 2 Patients, New Caledonia, 2014. *Emerging Infectious Diseases*. 2015 Feb;21(2):381–382. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313662/ [accessed 2016-04-04]. doi:10.3201/eid2102.141553.

[5] Petersen LR, Jamieson DJ, Powers AM, Honein MA. Zika Virus. *New England Journal of Medicine*. 2016 Apr;374(16):1552–1563. Available from: http://dx.doi.org/10.1056/NEJMra1602113 [accessed 2016-05-23]. doi:10.1056/NEJMra1602113.

[6] Musso D, Gubler DJ. Zika Virus. *Clinical Microbiology Reviews*. 2016 Jan;29(3):487–524. Available from: http://cmr.asm.org/content/29/3/487 [accessed 2016-05-12]. doi:10.1128/CMR.00072-15.

[7] Cao-Lormeau VM, Blake A, Mons S, Lastère S, Roche C, Vanhomwegen J, et al. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *The Lancet*. 2016 Feb;387(10027). Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673616005626 [accessed 2016-05-19]. doi:10.1016/S0140-6736(16)00562-6.

[8] Schuler-Faccini L, Ribeiro EM, Feitosa IML, Horovitz DDG, Cavalcanti DP, Pessoa A, et al. Possible Association Between Zika Virus Infection and Microcephaly — Brazil, 2015. *MMWR Morbidity and Mortality Weekly Report*. 2016 Jan;65(3):59–62. Available from: http://www.cdc.gov/mmwr/volumes/65/wr/mm6503e2.htm [accessed 2016-05-23]. doi:10.15585/mmwr.mm6503e2.

[9] Diekmann O, Heesterbeek JaP, Roberts MG. The construction of next-generation

matrices for compartmental epidemic models. *Journal of The Royal Society Interface*. 2010 Jun;7(47):873–885. Available from: http://rsif.royalsocietypublishing.org/content/7/47/873 [accessed 2016-05-12]. doi:10.1098/rsif.2009.0386.

[10] van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*. 2002 Dec;180:29–48.

[11] Andrieu C, Doucet A, Holenstein R. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010;72(3):269–342. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2009.00736.x/full [accessed 2015-02-12].

[12] Centre d'hygiène et de salubrité publique. Bulletin de la dengue et du zika en Polynésie française; 2014. Available from: http://www.hygiene-publique.gov.pf/IMG/pdf/bulletin_dengue_28-03-14.pdf [accessed 2016-05-12].

[13] Mallet HP, Vial AL, Musso D. Bilan de l'épidémie à virus Zika en Polynésie française. *Bulletin d'Informations Sanitaires Epidemiologiques et Statistiques*. 2015 May;(13). Available from: http://www.hygiene-publique.gov.pf/IMG/pdf/no13_-_mai_2015_-_zika.pdf [accessed 2016-05-12].

[14] Aubry M, Teissier A, Roche C, Teururai S, Desprès P, Mallet HP, et al.. Serosurvey of dengue, Zika and other mosquito-borne viruses in French Polynesia, abstr 765. Philadelphia, PA; 2015. Available from: http://www.abstractsonline.com/Plan/ViewAbstract.aspx?sKey=2e5199c4-aceb-4d61-8769-342586917c5a&cKey=74ffe328-3a83-4f28-84c6-5656971501c8&mKey=%7bAB652FDF-0111-45C7-A5E5-0BA9D4AF5E12%7d [accessed 2016-05-17].

[15] Direction des Affaires Sanitaires et Sociales. Situation sanitaire en Nouvelle Calédonie.

Les arboviroses : dengue, chikungunya, zika; 2014. Available from: http://www.dass.gouv.nc/portal/page/portal/dass/librairie/fichiers/32142267.PDF [accessed 2016-05-27].

[16] 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 [accessed 2015-03-09]. doi:10.1016/j.mbs.2013.10.007.

[17] 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). Available from: http://dx.plos.org/10.1371/journal.pcbi.1000898 [accessed 2015-01-27]. doi:10.1371/journal.pcbi.1000898.

[18] 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 [accessed 2015-09-08]. doi:10.1214/08-AOAS201.

[19] Anderson RM, May RM. Directly transmitted infections diseases: control by vaccination. *Science*. 1982 Feb;215(4536):1053–1060. Available from: http://science.sciencemag.org/content/215/4536/1053 [accessed 2016-05-23]. doi:10.1126/science.7063839.

[20] Wang L, Valderramos S, Wu A, Ouyang S, Li C, Brasil P, et al. From Mosquitos to Humans: Genetic Evolution of Zika Virus. *Cell Host & Microbe*. 2016 May;19(5):561–565. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1931312816301421 [accessed 2016-06-02]. doi:10.1016/j.chom.2016.04.006.

[21] Ledermann JP, Guillaumot L, Yug L, Saweyog SC, Tided M, Machieng P, et al. Aedes hensilli as a Potential Vector of Chikungunya and Zika Viruses. *PLOS Negl Trop Dis*.

2014 Oct;8(10):e3188. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003188 [accessed 2016-05-27]. doi:10.1371/journal.pntd.0003188.

[22] Kucharski AJ, Funk S, Eggo RM, Mallet HP, Edmunds WJ, Nilles EJ. Transmission Dynamics of Zika Virus in Island Populations: A Modelling Analysis of the 2013-14 French Polynesia Outbreak. *PLOS Negl Trop Dis*. 2016 May;10(5):e0004726. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004726 [accessed 2016-05-19]. doi:10.1371/journal.pntd.0004726.

[23] Funk S, Kucharski AJ, Camacho A, Eggo RM, Yakob L, Edmunds WJ. Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus; 2016. biorxiv;043265v3. Available from: http://biorxiv.org/lookup/doi/10.1101/043265 [accessed 2016-04-14].

[24] Nishiura H, Kinoshita R, Mizumoto K, Yasuda Y, Nah K. Transmission potential of Zika virus infection in the South Pacific. *International Journal of Infectious Diseases*. 2016 Apr;45:95–97. Available from: http://www.ijidonline.com/article/S1201971216000370/abstract [accessed 2016-04-27]. doi:10.1016/j.ijid.2016.02.017.

[25] Cazelles B, Hales S. Infectious Diseases, Climate Influences, and Nonstationarity. *PLOS Med*. 2006 Aug;3(8):e328. Available from: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0030328 [accessed 2016-05-24]. doi:10.1371/journal.pmed.0030328.

[26] Metcalf CJE, Farrar J, Cutts FT, Basta NE, Graham AL, Lessler J, et al. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *The Lancet*. 2016 Apr;Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673616301647 [accessed 2016-05-10]. doi:10.1016/S0140-6736(16)30164-7.

[27] Thompson RN, Gilligan CA, Cunniffe NJ. Detecting Presymptomatic

Infection Is Necessary to Forecast Major Epidemics in the Earliest Stages of Infectious Disease Outbreaks. *PLOS Comput Biol*. 2016 Apr;12(4):e1004836. Available from: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004836 [accessed 2016-05-12]. doi:10.1371/journal.pcbi.1004836.

[28] 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 [accessed 2016-05-24]. doi:10.1098/rsif.2012.0921.

[29] Smith DL, Perkins TA, Reiner RC, Barker CM, Niu T, Chaves LF, et al. Recasting the theory of mosquito-borne pathogen transmission dynamics and control. *Transactions of The Royal Society of Tropical Medicine and Hygiene*. 2014 Jan;108(4):185–197. Available from: http://trstmh.oxfordjournals.org/content/108/4/185 [accessed 2016-05-24]. doi:10.1093/trstmh/tru026.

[30] Halstead SB. Dengue Virus–Mosquito Interactions. *Annual Review of Entomology*. 2008;53(1):273–291. Available from: http://dx.doi.org/10.1146/annurev.ento.53.103106.093326 [accessed 2016-05-24]. doi:10.1146/annurev.ento.53.103106.093326.

[31] Bowman LR, Runge-Ranzinger S, McCall PJ. Assessing the Relationship between Vector Indices and Dengue Transmission: A Systematic Review of the Evidence. *PLOS Negl Trop Dis*. 2014 May;8(5):e2848. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002848 [accessed 2016-05-24]. doi:10.1371/journal.pntd.0002848.

[32] Harrington LC, Fleisher A, Ruiz-Moreno D, Vermeylen F, Wa CV, Poulson RL, et al. Heterogeneous feeding patterns of the dengue vector, Aedes aegypti, on

individual human hosts in rural Thailand. *PLoS neglected tropical diseases*. 2014 Aug;8(8):e3048. doi:10.1371/journal.pntd.0003048.

[33] Favier C, Degallier N, Rosa-Freitas MG, Boulanger JP, Costa Lima JR, Luitgards-Moura JF, et al. Early determination of the reproductive number for vector-borne diseases: the case of dengue in Brazil. *Tropical Medicine and International Health*. 2006 Mar;11(3):332–340. Available from: http://doi.wiley.com/10.1111/j.1365-3156.2006.01560.x. doi:10.1111/j.1365-3156.2006.01560.x.

[34] 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/. doi:10.1371/journal.pntd.0003719.

[35] Massad E, Burattini MN, Coutinho FAB, Lopez LF. Dengue and the risk of urban yellow fever reintroduction in Sao Paulo State, Brazil. *Revista de Saúde Pública*. 2003;37(4):477–484. Available from: http://www.scielo.br/scielo.php?pid=S0034-89102003000400013&script=sci_arttext&tlng=es.

[36] Aubry M, Finke J, Teissier A, Roche C, Broult J, Paulous S, et al. Seroprevalence of arboviruses among blood donors in French Polynesia, 2011–2013. *International Journal of Infectious Diseases*. 2015 Dec;41:11–12. Available from: http://www.sciencedirect.com/science/article/pii/S1201971215002398 [accessed 2016-03-07]. doi:10.1016/j.ijid.2015.10.005.

[37] Institut National de la Statistique et des Etudes Economiques. Populations légales au recensement de la population 2012 de Polynésie française;. Available from: http://www.insee.fr/fr/themes/detail.asp?ref_id=populegalescom&page=recensement/populegalescom/polynesie.htm [accessed 2016-05-12].

[38] Institut National de la Statistique et des Etudes Economiques. Populations légales

au recensement de la population 2014 de Nouvelle-Calédonie;. Available from: http://www.insee.fr/fr/themes/detail.asp?ref_id=populegalescom&page=recensement/populegalescom/nouvelle_caledonie.htm [accessed 2016-05-12].

[39] Roy M, Bouma MJ, Ionides EL, Dhiman RC, Pascual M. The Potential Elimination of Plasmodium vivax Malaria by Relapse Treatment: Insights from a Transmission Model and Surveillance Data from NW India. *PLOS Negl Trop Dis*. 2013 Jan;7(1):e1979. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0001979. doi:10.1371/journal.pntd.0001979.

[40] Lloyd AL. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings Biological Sciences / The Royal Society*. 2001 May;268(1470):985–993. doi:10.1098/rspb.2001.1599.

[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 [accessed 2015-03-27].

[42] Nishiura H, Mizumoto K, Villamil-Gómez WE, Rodríguez-Morales AJ. Preliminary estimation of the basic reproduction number of Zika virus infection during Colombia epidemic, 2015–2016. *Travel Medicine and Infectious Disease*. 2016 Apr;0(0). Available from: http://www.travelmedicinejournal.com/article/S1477893916300084/abstract [accessed 2016-04-27]. doi:10.1016/j.tmaid.2016.03.016.

[43] Bearcroft WGC. Zika virus infection experimentally induced in a human volunteer. *Transactions of The Royal Society of Tropical Medicine and Hygiene*. 1956 Jan;50(5):442–448. Available from: http://trstmh.oxfordjournals.org/content/50/5/442 [accessed 2016-05-12]. doi:10.1016/0035-9203(56)90091-8.

[44] Lessler J, Ott C, Carcelen A, Konikoff J, Williamson J, Bi Q, et al. Times to key events in the course of Zika infection and their implications: a systematic review

and pooled analysis. *Bull World Health Organ [Internet]*. 2016;Available from: http://www.who.int/entity/bulletin/online_first/BLT.16.174540.pdf?ua=1 [accessed 2016-05-12].

[45] Hayes EB. Zika Virus Outside Africa. *Emerging Infectious Diseases*. 2009 Sep;15(9):1347–1350. Available from: http://wwwnc.cdc.gov/eid/article/15/9/09-0442_article.htm [accessed 2016-02-22]. doi:10.3201/eid1509.090442.

[46] Chouin-Carneiro T, Vega-Rua A, Vazeille M, Yebakima A, Girod R, Goindin D, et al. Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. *PLOS Negl Trop Dis*. 2016 Mar;10(3):e0004543. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004543 [accessed 2016-03-07]. doi:10.1371/journal.pntd.0004543.

[47] Brady OJ, Johansson MA, Guerra CA, Bhatt S, Golding N, Pigott DM, et al. Modelling adult Aedes aegypti and Aedes albopictus survival at different temperatures in laboratory and field settings. *Parasites & Vectors*. 2013;6:351. doi:10.1186/1756-3305-6-351.

[48] 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 [accessed 2016-05-13]. doi:10.1371/journal.pone.0089783.

[49] Direction des Affaires Sanitaires et Sociales. Point sur la situation épidémiologique et les mesures sanitaires; 2014. Available from: http://www.gouv.nc/portal/pls/portal/docs/1/28032255.PDF [accessed 2016-05-12].

[50] Doucet A, Freitas N, Gordon N, editors. Sequential Monte Carlo Methods in Practice. New York, NY: Springer New York; 2001. Available from: http://link.springer.com/10.1007/978-1-4757-3437-9 [accessed 2016-05-20].

[51] Doucet A, Johansen AM. A tutorial on particle filtering and smoothing: fifteen years later; 2011.

[52] Roberts GO, Rosenthal JS. Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics*. 2009 Jan;18(2):349–367. Available from: http://dx.doi.org/10.1198/jcgs.2009.06134 [accessed 2015-05-18]. doi:10.1198/jcgs.2009.06134.

[53] Dietz K. The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research*. 1993;2(1):23–41. Available from: http://smm.sagepub.com/content/2/1/23.short.

[54] Blower SM, Dowlatabadi H. Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: An HIV Model, as an Example. *International Statistical Review / Revue Internationale de Statistique*. 1994;62(2):229–243. Available from: http://www.jstor.org/stable/1403510. doi:10.2307/1403510.

[55] Carnell R, Carnell MR, RUnit, Suggests. Package 'lhs'. . 2016;Available from: http://cran.stat.auckland.ac.nz/web/packages/lhs/lhs.pdf.

[56] Pujol G, Iooss B, Janon A. Package 'sensitivity'; 2016. Available from: https://cran.r-project.org/web/packages/sensitivity/sensitivity.pdf.