Modelling seasonality of Lassa fever in Nigeria

James McKendrick[1,*], Warren Tennant[2], Michael J. Tildesley[2]

1 MathSys, Mathematical Institute, Zeeman Building, University of Warwick, Coventry, United Kingdom
2 Zeeman Institute: SBIDER, School of Life Sciences and Mathematics Institute, University of Warwick, Coventry, United Kingdom

* j.mckendrick@warwick.ac.uk

Abstract

Lassa fever (Lf) is a viral haemorrhagic disease endemic to West Africa and is caused by the *Lassa mammarenavirus*. The rodent *Mastomys natalensis* serves as the primary reservoir and its ecology and behaviour have been linked to the distinct spatial and temporal patterns in the incidence of Lf. Nigeria has experienced an unprecedented epidemic that lasted from January until April of 2018, which has been followed by subsequent epidemics of Lf in the same period every year since. While previous research has modelled the case seasonality within Nigeria, this did not capture the seasonal variation in the reproduction of the zoonotic reservoir and its effect on case numbers. To this end, we introduce an approximate Bayesian computation scheme to fit our model to the case data from 2018–2020 supplied by the NCDC. In this study we used a periodically forced seasonal nonautonomous system of ordinary differential equations as a vector model to demonstrate that the population dynamics of the rodent reservoir may be responsible for the spikes in the number of observed cases in humans. The results show that in December through to March, spillover from the zoonotic reservoir drastically increases and spreads the virus to the people of Nigeria. Therefore to effectively combat Lf, attention and efforts should be concentrated during this period.

Author summary

Lassa fever is a viral disease prevalent in West Africa, with *Mastomys natalensis* serving as the primary reservoir. In Nigeria, annual outbreaks occur from December to March. Using a novel model and data from 2018-2020, we demonstrate that the population dynamics of the reservoir contribute to spikes in human cases. Specifically, spillover transmission increases drastically during this period, highlighting the need for concentrated efforts and interventions. Understanding the seasonal dynamics of the reservoir is crucial for effective Lassa fever control and prevention strategies in Nigeria.

Introduction

Lassa fever (Lf) is a viral zoonotic disease, caused by the *Lassa mammarenavirus* (LASV), that is endemic to West African countries such as Nigeria, Sierra Leone and Guinea [1, 2]. Lf has a natural reservoir in the rodent *Mastomys natalensis* in which the virus persists and crossover events to humans occur [3, 4, 5]. The disease, which was first described after two nurses contracted the disease in a hospital in Jos, Nigeria, in

1969, has since been identified as a significant risk to health in West Africa with 300,000–500,000 cases per year resulting in approximately 5,000 deaths annually [3, 6]. Those infected with LASF typically experience acute symptoms of headaches, sore throat, muscle pain, vomiting and diarrhoea, and in severe cases bleeding from the mouth, nose, vagina or gastrointestinal tract [1]. The risk that Lf poses to public health will only increase without widespread intervention and a viable vaccine as growth in inter-border traffic and international travel increases the likelihood of introducing the virus to other regions within and outside of the African continent [7, 8].

In recent years, Nigeria has experienced epidemics with peak incidence occurring between December and April in the years since 2017, just after the rainy season ends in October. Both the number of cases and the exposure rate of Lf increases in certain periods of the year and have been correlated with rainfall patterns [9, 10, 11]. This may be because the reproduction of *Mastomys* is greatest just after the rainy season, which results in an increase in the size of the rat population and in the spread of LASV from infected to susceptible rats[12, 13, 14, 15]. The ecological dynamics of *M. natalensis* are relevant to Lf in humans because the majority of infections (80%) are suspected to be spillover events as opposed to human-to-human transmissions [16]. The importance of rat contact with humans is illustrated by data reported by Tobin et al. (2015) for Edo state, Nigeria, which recorded 32.4%(385/1189) of the confirmed national cases of Lf in 2020; 96.1% houses had found the multimammate rat within them in the past 6 months and 58.2% of the resident were seropositive (i.e. tested positive for Lf-specific antibodies) [17]. Hence the relationship between the presence and behaviour of *M. natalensis* and the prevalence of Lf is critical to understanding and predicting future outbreaks.

Mathematical modelling studies of Lf are rare compared with other diseases, despite the inclusion of Lf in the World Health Organization's Blueprint list of diseases to be prioritized for research and development [18]. Published models for recent outbreaks in Nigeria can incorporate the population dynamics of the disease reservoir and highlight areas and periods of the year at high risk of transmission [19, 20, 21]. Akhmetzhanov et al. (2019) used suspected case data and a rodent model to inform a time-dependent exposure rate of Lf to susceptible people in Nigeria, but did not include explicit modelling of human infection dynamics [9]. To date, the mechanistic models describing Lf epidemics in Nigeria lacked the focus on time-dependent parameters relating to the rodent population dynamics to explain the relationship between those dynamics and the seasonality of outbreaks.

In this study, we developed an epidemiological model to describe the temporal dynamics of Lf within Nigeria in both human hosts and rodent vectors incorporating seasonal variations in rodent population dynamics. The model was then fitted to confirmed case data from the Nigerian Centre for Disease Control (NCDC) at the national scale in Nigeria using Approximate Bayesian Computation (ABC). This approach allowed us to demonstrate how the annual fluctuations of the rodent population translated into the seasonal outbreaks of Lf cases in human hosts.

We believe that providing an epidemiological model of the dynamics of Lf in humans and the rodent reservoir within Nigeria as a whole that fits to confirmed cases supports the hypothesis that the fluctuations of the rodent population strongly influence the cases observed. In order to inform future public health efforts, it is important to better understand the role of the disease reservoir in the spatio-temporal profile of infections.

Materials and methods

The weekly situation reports on Lf produced by the NCDC provided a stream of publicly available data that were discussed in regular direct communications with NCDC representatives. The data used in this study was the set of dated Lf cases collected by

the NCDC's surveillance network for Lf between 7$^{\text{th}}$ January 2018 until 12$^{\text{th}}$ July 2020, which covers the 2018, 2019 and 2020 epidemics [22]. Lf cases were categorised as either suspected, confirmed and probable. Confirmed cases were those which had a positive result for IgM antibody, PCR or virus isolation. Suspected cases were any individual experiencing symptoms such as fever, sore throat, vomiting, diarrhoea. Additionally they also met one of the following criteria: if they had a history of contact with either 1) excreta or urine of rodents, 2) with a probable or confirmed Lf case recently, or 3) any person with inexplicable bleeding/haemorrhagia. Due to the uncertainty that would result from use of unconfirmed data we used only confirmed cases.

Model

In order to describe the confirmed Lf cases in Nigeria, we constructed a vector-host model in which the dynamics of the rodent *M. natalensis* are modelled explicitly in addition to the dynamics of human transmission and disease progression (Fig 1 and equations 1–3).

The human population was split into susceptible, exposed, asymptomatic or infected, and recovered $(S_h, E_h, A_h, I_h, R_h)$. The infection pathway for humans was described in the model as follows: Susceptible people acquire infections from the pool of all infectious individuals. This is at a rate of λ_h. Contact with an infectious host, whether it is a human or vector rodent, transmits the disease to the susceptible individual and then they become exposed. Exposed humans become infectious after a period of $1/\nu$ days, and thus exposed individuals will leave the compartment at a rate of ν. Those who were exposed to the virus will then become Asymptomatic with a probability of p and Infected with a probability of $1-p$. Infected humans are assumed to experience more severe symptoms and be recorded as cases in data; they also experience an infection induced mortality rate μ_{h_I}. Infected individuals are assumed to be detectable. Infected individuals will recover with an average recovery period of $1/\gamma_h$ days.

All humans are assumed to have an average life expectancy of $1/\mu_h$ days regardless of their compartment. That is, the number of individuals in each compartment decrease at a rate of μ_h per capita. Humans are assumed to have a constant birth rate, B_h, and newborns are assumed to be fully susceptible to Lf.

The vector population of rodents was split into susceptible, infected, and recovered (S_r, I_r, R_r). Rats are born susceptible at a time-dependent per capita rate of $B(t)$. The rat infections follow a more simple pathway than that found in humans: Susceptible rats come into contact with infected rats, whereby the disease is transmitted to the susceptible individual. This occurs at a rate of λ_r. The infected rats do not experience increased mortality and thus the rats have a constant mortality rate of μ_r. Infectious individuals will recover with an average recovery period of $1/\gamma_r$.

M. natalensis has repeatedly been observed to have seasonal breeding habits over different areas of Africa [12, 13, 14]. Therefore the recruitment rate per capita of the rats $B(t)$ was chosen as

$$B(t) = k \exp\left\{-s \cos\left(\left(\pi\left(\frac{t}{365} - \phi\right)\right)\right)^2\right\} \tag{1}$$

where k is the magnitude of the function; s is a shape parameter denoting how long the period of low reproduction rates lasts for, a smaller s meaning a close to constant recruitment rate over the year whereas a larger s would equate to a long low period then a sharper change to a high period; and ϕ is the point in the year where the reproduction of the rats is at its minimum. k and s are of positive value, and ϕ is between 0 and 1. Once parameters s and the natural mortality rate of the rats, μ_r, have been fixed, k may be scaled to keep the reservoir population constant, year-on-year [23].

A population whose yearly dynamics are similar is a better representation of a species that is endemic to the environment and although fluctuations will happen, the data to accurately model the population does not exist.

The rate at which susceptibles become infected, otherwise known as the force of infection, is denoted by λ_h and λ_r for humans and rats respectively. In humans, this is defined as a linear combination of the possible contact routes with LASV carriers. To represent the difference in transmissibility of Lf in rats and humans, the contact rates β_{rh} and β_{hh} are separated. Explicitly, these parameters are the rate of successful transmissions from rats to humans per infected rat and humans to humans per infected human respectively. In addition, we have assumed a density dependent contact rate between susceptible humans and the pool of infectious individuals, which means that contacts will occur at an invariable rate irrespective of the size of the human population.

With the transmission rates between compartments defined, the susceptible compartments experience a force of infection from the infectious agents. The total force of infection per susceptible is therefore the sum linear combination of the number of infectious agents that can infect that susceptible host multiplied by the appropriate transmission rate. Therefore the force of infection per individual human and rat, respectively denoted λ_h and λ_r, are as follows:

$$\begin{aligned} \lambda_h &= \frac{\beta_{rh} I_r + \beta_{hh} \left(A_h + I_h\right)}{N_h} \\ \lambda_r &= \frac{\beta_{rr} I_r}{N_r} \end{aligned} \tag{2}$$

$$\begin{aligned} \frac{\mathrm{d}S_r}{\mathrm{d}t} &= B(t)N_r - \lambda_r S_r - \mu_r S_r \\ \frac{\mathrm{d}I_r}{\mathrm{d}t} &= \frac{\beta_{rr} S_r I_r}{N_r} - (\gamma_r + \mu_r) I_r \\ \frac{\mathrm{d}R_r}{\mathrm{d}t} &= \gamma_r I_r - \mu_r R_r \\ \frac{\mathrm{d}S_h}{\mathrm{d}t} &= B_h N_h - \lambda_h S_h - \mu_h S_h \\ \frac{\mathrm{d}E_h}{\mathrm{d}t} &= (\beta_{rh} I_r + \beta_{hh}(A_h + I_h))\frac{S_h}{N_h} - (\nu + \mu_h) E_h \\ \frac{\mathrm{d}A_h}{\mathrm{d}t} &= p\nu E_h - \mu_h A_h \\ \frac{\mathrm{d}I_h}{\mathrm{d}t} &= (1-p)\nu E_h - (\gamma_h + \mu_{h_I} + \mu_h) I_h \\ \frac{\mathrm{d}R_h}{\mathrm{d}t} &= \gamma_h I_h - \mu_h R_h \end{aligned} \tag{3}$$

Fig 1. Model flowchart of the transmission and population dynamics of the system of equations 3. Blue solid arrows denote recruitment. Black solid arrows denote progression of the disease. Red dashed arrows denote disease transmission. Purple solid arrows denote mortalities. Parameters are detailed in full in table 1 where λ_h and λ_r are defined in equations 2 (i) and (ii) respectively, and $B(t)$ is defined in equation 1.

Basic reproduction numbers

The basic reproduction ratio, R_0, is the expected number of secondary infections caused by a single infectious agent in an otherwise susceptible population. Over time, the conditions which the infectious agent inhabit change and thus we also calculate the effective reproduction rate, which shows the expected number of secondary infections from one infectious agent in the population (which may have other infectious agents) at a specified time t.

Based on the next generation method for deriving the reproductive ratio, R_0, from Diekmann et al (1990) and the particular method used in van den Driessche and Watmough's work (2002) we produce the effective reproductions rates between species in equation set 4 with derivations in S1 appendix (Section 1.2) [24, 25].

$$R^{rr}(t) = \frac{\beta_{rr}}{(\gamma_r + \mu_r)} \frac{S_r}{N_r} \tag{4}$$

$$R^{rh}(t) = \frac{\beta_{rh}}{(\gamma_r + \mu_r)} \frac{S_h}{N_h} \tag{5}$$

$$R^{hh}(t) = \left(\frac{p\nu\beta_{hh}}{(\nu + \mu_h)(\mu_h + \gamma_h)} + \frac{(1-p)\nu\beta_{hh}}{(\nu + \mu_h)(\mu_{h_I} + \mu_h + \gamma_h)} \right) \frac{S_h}{N_h} \tag{6}$$

Parameter Selection

In table 1 we show the choices for the model parameters that are described in this subsection.

Outbreak events within Nigeria have a seasonal pattern, thus each epidemic should not be analysed in isolation as this would neglect the effect of seasonal dynamics. Therefore the data fitting takes place over a longer time period and the population dynamics of the people of Nigeria should be taken into account. The natural death rate of humans μ_h is estimated to be $\frac{1}{54\times365}$ day^{-1} since 54 years was the average life expectancy to 2 significant figures for Nigerians given by the World Bank for 2018. The human birth rate B_h was approximated as 1.2×10^{-4}day^{-1}. Nigeria has experienced close to exponential growth rate in recent years and if it is assumed that the population growth of Nigeria has been stable over this time period then we may assume that $\frac{\mathrm{d}N}{\mathrm{d}t} = (B_h - \mu_h)N_h$, where B_h and μ_h are constant. Therefore $N(t) = N(0)e^{(B_h-\mu_h)t}$. We then obtain $B_h = \mu_h + \frac{1}{T}\log\frac{N(T)}{N(0)}$. The growth of Nigeria from 2015 to 2019 was 181.1 million to 201.0 million to 4 significant figures which gives the growth rate of Nigeria to be 7.14×10^{-5} per capita per day hence giving $B_h = \mu_h + 7.14 \times 10^{-5}$ [26].

The incubation period was assumed to be 14 days. There is a wide range for the incubation period and is reported to be around 2 days to 3 weeks. For simplification this assumption was made to be approximately 14 days and thus $\nu = 1/14$. The rate of recovery for humans γ_h was 0.1 day^{-1}. Ranges for the time to recover are broad, between 2 and 21 days, so similarly to ν a value of 0.1 was assumed. We assume that the probability of becoming asymptomatic $p = 0.8$ [1].

The infection induced mortality rate $\mu_{hI} = 0.195\gamma = 0.0195$ day^{-1}. The proportion of those that have died, retrospectively, during the outbreak period considered is 196/1006. Therefore the case fatality rate (CFR) is 19.5% to 3 significant figures [22]. To give a rate of mortality the CFR is multiplied by the recovery rate to approximate the rate that individuals die from Lf before they recover[27].

The initial number of infected humans $I_h(0)$ is 2 since there were two recorded confirmed cases in the week commencing 01/01/2018. Therefore $A_h(0) = 8$ to maintain the ratio between asymptomatic and symptomatic infected persons to 1:4. $E_h(0)$ is 5 times that of the number of cases reported the week after the data being used starts.

Parameter	Value/Prior	Biological Description
B_h	$1.2 \times 10^{-4}\text{day}^{-1}$	Human birth rate
μ_h	$\frac{1}{54\times365}$ day^{-1}	Human natural mortality rate
ν	0.1 day^{-1}	Reciprocal of incubation period
γ_h	0.1 day^{-1}	Human recovery rate
p	0.8	Probability of an infectious human being asymptomat
μ_{h_I}	0.0195 day^{-1}	Infection induced mortality in humans
γ_r	1/90 day^{-1}	Recovery rate for rats
μ_r	1/500 day^{-1}	Natural mortality rate for rats
β_{rr}	$LogNormal(-1.03, 1)$	Rat-to-rat transmission rate
β_{rh}	$LogNormal(-7.77, 0.5)$	Rat-to-human transmission rate
β_{hh}	$LogNormal(-2.35, 0.5)$	Human-to-human transmission rate
ϕ	$Uniform(0, 1)$	Time of minimum reproduction for rats
s	$LogNormal(3, 1)$	Shape parameter for reproduction function for rats
$N_h(0)$	2×10^8	Initial number of humans
$S_h(0)$	10^5	Initial number of susceptible humans
$E_h(0)$	30	Initial number of exposed humans
$A_h(0)$	8	Initial number of asymptomatic humans
$I_h(0)$	2	Initial number of infected humans
$N_r(0)$	10^6	Initial number of rats
$S_r(0)$	$N_r(0)/R_0^{rr}$	Initial number of susceptible rats
$I_h(0)$	$N_r(0)\frac{\mu_r}{\beta_{rr}}(R_0^{rr} - 1)$	Initial number of infected rats

Table 1. Fixed parameters and parameters to be estimated in the Lassa fever model. The parameters of interest were inferred using algorithm 1 in section 2.4.

This is done to represent that those who show symptoms were likely exposed the week before and that 20% of the exposed will go on to show symptoms. The initial total number of people living in Nigeria was assumed to be 2×10^8. The initial number of susceptible $S_h(0)$ is therefore the remaining number of humans, 2×10^8.

The mortality rate of rats is assumed as $\mu_r = 1/500$ per day since this value has been previously used as the baseline value in previous works [20, 28]. $\gamma_r = 1/90$ per day [29]. For the population size of the zoonotic reservoir we tested four initial values for size, $N_r(0) \in \{1, 10^6, 5 \times 10^6, 2.5 \times 10^7\}$. The number of initially susceptible rats was $S_r(0) = N_r(0)/R_0^{rr}$. This is necessarily bounded by $N_r(0)$ above and below by 0 since $beta_r r$ may be sampled such that $R_0^{rr} < 1$. $I_r(0)$, the total initial rat population, was taken from an array of values to test for the impact on the system, $I_r(0) \in \{\frac{N_r(0)}{2}, \frac{N_r(0)}{10}, \frac{N_r(0)}{100}, \mu_r N_r(0)(R^r r_0 - 1)/\beta_{rr}\}$. The later value was chosen as it is the endemic equilibrium of an SIR system with a recruitment/birth function equal to μ_r. The remaining rats were assumed to be recovered.

The remaining parameters, $\phi, s, \beta_{rr}, \beta_{rh}$ and β_{hh} will be estimated within the fitting scheme in the following section.

Fitting and Data

Bayesian estimation techniques involve a suite of statistical inference methods based on the idea that after specifying a prior assumption upon the parameters being investigated, the prior is updated with the introduction of more information from the observed data. Following Bayes' theorem, the posterior distribution of parameters is obtained by combining the prior beliefs (prior distribution) with the evidence of the data which usually comes in the form of the likelihood function [30].

In the absence of a likelihood function, which may arise because the function is

intractable or computationally expensive, Approximate Bayesian Computation (ABC) is a robust method that can be used. ABC can be summarised as a family of techniques where parameters are sampled, in various different ways that are dependent on the specific scheme, and then accepted or rejected if the simulated data given the parameters are sufficiently close to the observed data [31]. ABC schemes are becoming increasing popular due to their relative ease of use.

In this paper we use a modified Approximate Bayesian Computation Sequential Monte Carlo scheme (ABC SMC) to fit our parameters (see Algorithm 1 for psuedocode description of scheme used) since the scheme has been shown to be reliable and converges faster than some of the more primitive schemes [32]. The ABC SMC schema iterates a population of parameter particles over T generations with decreasing tolerances, $\{\varepsilon_i\}_{i=1}^{T}$, allowed between the data, y^*, and the data simulated from the model with the particle θ, y_θ. This converges to the desired approximate posterior distribution as the distributions of the parameters are sequentially improved upon [33].

ABC SMC fits a model M with unknown model parameters θ to data. The standard algorithm requires one to specify a decreasing sequence of thresholds $\varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_T$ for the T generations. When starting, $t = 1$, parameters are sampled from prior distributions, $\pi(\theta)$. For each subsequent generation $t = 2, \cdots, T$ parameters will be sampled from a perturbation kernel, $q_t(\theta|\theta_{t-1}^{(i)})$, based on a sampled particle accepted in generation $t-1$. The model is then simulated with the sampled parameter particle and using a chosen distance metric to compare the data and the simulated data, the parameters are accepted if the error calculated is smaller than the given tolerance for that generation, i.e. $d(y^*, y_\theta) \leq \varepsilon_t$.

When implementing the SMC algorithm, instead of manually defining the sequence $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T$, which is often done by manually calibrating the tolerances after some initial test runs, we instead initialise our algorithm with an integer, K, and a proportion, Q. If the desired number of parameter particles is N then we initialise by sampling KN particles from the prior distributions and reject all the N particles with the smallest errors. This serves as our first generation of sampling. For subsequent generations we have set a desired quantile, Q, where the particle that is the Q^{th} quantile has its error between the data and model set as the tolerance for the next generation. That is, if the set of parameter particles for generation $g-1$ is $\{\theta_i^{g-1}\}_{i=1}^{N}$ and is ordered with respect to $d(y^*, y_\theta)$ then for generation g the tolerance $\varepsilon_g = d(y^*, y_{\theta_Q^{g-1}})$. Furthermore the perturbation kernel that we use is a multivariate Gaussian with variance equal to twice that of the co-variance between the previous generation's particles: $q_t(\theta|\theta_{t-1}^{(i)}) = \mathcal{N}(\theta|\theta_{t-1}^{(i)}, \Sigma_{t-1})$, where i is sampled using the weightings generated.

Prior Distributions

For the parameters that we are unable to calculate, we fit with the scheme from section 2.4 specifically described as Algorithm 1. In this section we explain the choices for the prior distributions from our prior knowledge on the model and disease epidemiology.

ϕ is sampled from a uniform distribution between 0 and 1 because this does not give any preference to any period of the year over another. s is sampled from a Log-normal distribution, Lognormal(μ, σ^2), where $\mu = 3$ and $\sigma = 1$ as this allows the parameter samples to be varied and comparable to the range used in Peel et al (2014) [23]. $\beta_{rr} \sim$ Lognormal (μ, σ^2). Since the zoonotic system is assumed to be in endemic equilibrium, R_0^{rr} is assumed to be greater than 1. We therefore set μ so that the mode of the distribution, $m_1 = \exp\big(\mu - \sigma^2\big)$ where $\sigma = 1$, would equate to $R_0^{rr} = \frac{m_1}{\mu_r + \gamma_r} = 10$, with a prior flexible enough to sample through parameter space for a variety of values. Therefore $\mu = \log(10(\gamma_r + \mu_r)) + \sigma^2 = -1.03$ to 3 s.f. As our system is not at the point of introduction of Lf, rather it is close to endemic equilibrium, the rate at which new

Input: N, number of particles per generation

K, multiples of N for the initial sampling to determine the sequence of tolerances ε_i

Q, Quantile between 0 and 1 to select the next tolerance from the distribution of tolerances from the previous generation

$\pi(\theta)$, Prior distribution for the tested variables

$q_t(\theta|\theta_{t-1}^{(i)})$, Method of perturbation to generate samples of particles for generations $t = 2, \cdots, T$

y_0, Data and method of determining closeness to simulated data $d(\cdot, \cdot)$

Model $M(\theta)$

Output: $\{\theta_i^T\}_{i=1}^N$, the accepted parameters from generation T

; `// Run the first generation`

for $i = 1$ *to* KN **do**

Generate θ^i from the prior $p(\theta)$ Generate data y_{θ^i} from the model $M(\theta_i)$ Calculate $d_i = d(y_{\theta^i}, y^*)$

end

Sort initial KN particles by their distance entries d_i;

Set $\{\theta_i^1\}_{i=1}^N$ to be the N best of the KN particles, (retroactively making $\varepsilon_1 = d_N$)

; `// Run subsequent generations`

for $t = 2$ *to* T **do**

Calculate weights $\omega_1^{(i)} \longleftarrow 1/N$ Set next tolerance with Q by letting $\varepsilon_2 = d_{floor(QN)}$ **while** $i =\leqslant N$ *do* **do**

Draw θ^* from among θ_{t-1} with probabilities ω_{t-1} Generate θ from $q_t(\theta|\theta_{t-1}^{(i)}) = \mathcal{N}(\theta^*, \Sigma_{t-1})$ where Σ_{t-1} is the covariance of the previous generation of particles Generate y_θ from the simulator $d(y_\theta, y^*) \leq \varepsilon_t$ $\theta_t^{(i)} \longleftarrow \theta$ $\omega_t^{(i)} \longleftarrow \frac{p(\theta)}{(\Sigma_{k=1}^N \omega_{t-1}^{(k)} calN(\theta|\theta_{t-1}^{(k)}, t-1))}$

end

$\varepsilon_{t+1} = d_{floor(QN)}$ where $floor(x)$ is the greatest integer less than x

end

Algorithm 1: Psuedocode of modified SMC ABC. This was used to fit the model in section 2.1. Instead of using an arbitrary sequence of tolerances, the tolerances are calculated from the errors produced in the previous generation. For the first generation, the algorithm runs a multiple, K, of the number of desired particles, N, and then accepts the best N.

infections are acquired by humans from rats will be proportional to N_r and so to factor this in we set the Log-normal prior distribution for β_{rh} to have a mode m_2 to be such that $R_0^{rh} = \frac{m_2}{\mu_r+\gamma_r} = \frac{10A}{N_r(0)}$ where $A = 2,500$. Thus the prior distribution has $\mu = \log(10(\gamma_r + \mu_r)) + \sigma^2 + \log(A/N_r(0)) = -7.77$ to 3 s.f. Since human-to-human transmissions are unlikely outside of nosocomial settings, we set $\sigma = 0.5$ and $\beta_{hh} \sim$ Lognormal(μ, σ^2) so that the mode m_3 when equated to β_{hh} and inserted into R_0^{hh} would give $R_0^{hh} = 1$. Therefore $\mu = \log(\psi) + \sigma^2 = -2.35$ to 3 s.f. where

$$\frac{1}{\psi} = \frac{p\nu}{(\nu+\mu_h)(\mu_h+\gamma_h)} + \frac{(1-p)\nu}{(\nu+\mu_h)(\mu_{h_I}+\mu_h+\gamma_h)} \tag{7}$$

As previously stated, the data used here are taken directly from our communications with the NCDC. The data points used are laboratory confirmed cases aggregated by week from the 12$^{\text{th}}$ of December 2018 until the 4$^{\text{th}}$ of April 2020.

Implementation

The fitting algorithm and all models were implemented in MATLAB, with the models using the ODE45 solver for numerical integration of the ODEs.

The algorithm ran for $T = 15$ generations with each generation consisting of 2500 parameter particles. For the first generation $K = 10$ multiples of 2500 particles were ran and then the $1/K$ quantile with the least error was accepted. Subsequent tolerances for each generation were set with the quantile $Q = 1/6$, where errors from the previous generation would generate the next tolerance value.

Results

We applied an ABC fitting scheme (Algorithm 1) to the model detailed in Materials and methods (Fig 1 and eq. 3) to the confirmed Lassa fever cases in Nigeria. Notably, our fitted model successfully replicated the observed seasonal trends in the data, as demonstrated in Figure 2. The posterior distributions of parameters and vector dynamics (Fig 3 and 4, respectively) reveal the seasonal nature of the epidemics and how the population dynamics of the primary reservoir, *M. natalensis*, affect the number of observed cases in Nigeria.

Model Evaluation

Fig 2. The epidemiological model captured 3 consecutive Lf epidemics in Nigeria. The simulated cases compared with the observed data. In orange is the entire range of values I_h takes in the final generation at each time point; the median value in blue. Confirmed case data for Nigeria are in black. The model replicates the sharp increase in case incidences occurring at the start of the year for 3 years.

The model fit resulted in a final generation whose simulations of the number of symptom-presenting humans, I_h, can be seen in Fig 2. The entire range of values that I_h takes in the final generation at each time point is in light orange and the median value in blue. Overlaid in black is the confirmed case data for Nigeria. This shows that the model can replicate the year-on-year trend and that the seasonal epidemics in Nigeria can be explained by the vector population dynamics.

The number of infected rats increases drastically in December when the pool of available susceptibles grows, thereby increasing the spillover rate to humans (Fig 4). As the number of rats in contact with humans decreases over the year due to natural

Fig 3. The marginal posterior distributions of the final set of accepted particles from fitting. Figure 3 top left the shape parameter of the rodent recruitment function, s. Figure 3 top right the rodent-to-rodent transmission rate β_{rr}. Figure 3 mid left the human-to-human transmission rate β_{hh}. Figure 3 mid right the rodent-to-human transmission rate β_{rh}. Figure 3 bottom left the date of minimum rat reproduction ϕ.

mortality and recovery the number of spillover events decreases. This process starts again just before the next epidemic and continues cyclically.

Marginal Posterior Parameter distributions

The recruitment rate of the vector *M. natalensis* is determined by the shape parameter, s fig 3 top left, and the date of the minimum rate, ϕ fig 3 bottom left, is focused and seasonal. Since the posterior of s has increased its median substantially (6.08×10^2), this results in a large ratio between the reproduction rate's lowest and highest value and thus the rat population experiences an influx of susceptible rats at ϕ+6 months in early December. The distribution for ϕ is concentrated around a median of the 7^{th} of June with a 90% credible interval of the 4^{th} – 11^{th} of June (see S1 appendix table 2 for unconverted values). This is observed in the rodent dynamics in fig 4 where the number of susceptible rats increases rapidly.

The time-varying reproduction rate for rat-to-rat transmission crosses the threshold of $R_t^{rr} \geq 1$ in early December (fig 5 (a)) and causes the number of infectious rats to increase similarly. This then spills over to humans and causes the spike in incidence data that is observed in Nigeria between January and March.

The transmission parameters, fig 3 top right and middle, are such that human infections are predominantly the result of a spillover event from the zoonotic reservoir. Despite rat-to-human transmission, β_{rh}, low estimation as seen in S1 appendix table 2 with the expected number of secondary human infections per infected rat in a completely susceptible population, R_0^{rr}, having a median of 4.70×10^{-3}, the number of infected rats during December through to May is sufficient to cause a significant number of infections in humans. Moreover, the proportion of humans infected by rats is estimated to be 97.42% – 99.72% (90% credible interval) with a median of 98.51%.

Biological Implications

In our study, we explicitly modelled the vector dynamics allowing us to investigate biological implications. The seasonal recruitment of the natural reservoir as seen in Fig 4 is a crucial aspect of the disease dynamics and serves as a key driver of the transmission cycle. We can infer from these results that *M. natalensis* recruitment, which may be an combination of birth and migration, influence the occurrence of spillover events to humans in Nigeria, with rat-to-rat infection peaking just before the epidemics observed in Nigeria (Figs 5 (a) and 5 (b)). This emphasizes the importance of understanding the ecological and reproductive dynamics of the reservoir species, as it directly impacts the risk of disease transmission to human populations.

Sensitivity Analysis

We conducted a sensitivity analysis (see S1 appendix 3.1) to assess the impact of varying the assumed population size of the zoonotic reservoir on the predicted number of reported cases. We conducted the test with populations of sizes 10^6, 5×10^6 and 2.5×10^7 where each was tested with $1/2, 1/20$ and $1/100$ of the population as being

Fig 4. Underlying vector dynamics reveal high-risk period of spill-over transmission for Nigerians. The figure showcases the evolution of *M. natalensis* compartments throughout the observed period, simulated using the parameters derived from the final generation of accepted values. The median value is represented by the dashed line, while the colored area illustrates the range. Susceptible rats are depicted in red, infected rats in green, and recovered rats in blue. Notably, the recruitment of susceptible rats progressively rises, providing impetus for the growth of infected rats, reaching its peak in late December. Consequently, this surge in infected rats leads to spillover infections in humans.

Fig 5. The range and median of the effective reproduction rate for rat-to-rat transmission $R^{rr}(t)$ and when the threshold for $R^{rr}(t) \geq 1$ is met. In Fig 5 (a) $R^{rr}(t)$, median dashed-line and range in coloured block, exhibits a sharp increase towards the end of the year, foreshadowing the subsequent outbreaks in the following months. To maintain clarity, the data is limited to the years 2019 and 2020, as no complete earlier records are available. Figure 5 (b) showcases a box diagram illustrating the time of year when $R^{rr}(t)$ exceeds the threshold of 1, denoting high transmission. The bottom panel of Fig 5 (b) captures the onset of the high transmission period, while the upper panel displays its conclusion. The intermediate phase witnesses a rapid shift in reservoir dynamics, leading to an escalation in the number of infected vectors

initially infected. We also tested with the original assumption of $I_r(0) = N_r(0)\frac{\mu_r}{\beta_{rr}}(R_0^{rr} - 1)$. We used the unaltered SMC algorithm with the tolerances set to be those calculated in the original run as seen in S1 appendix Table 1.

The results showed that the size of the zoonotic reservoir can be balanced by the transmission rate β_{rh} and it changes with the inverse change of the population (S1 appendix Fig 1 and 2). I.e. 5 times more total rats meant that the posterior distribution of β_{rh} was approximately a fifth of the original result as seen in 3 middle right.

The initial number of infected rats changed the beginning of the simulated case data but the fitted model was still able to capture the peaks of the epidemics in 2019 and 2020. The different initialisations of the case scenarios showed that the model was not flexible enough to fit to both the later epidemics and the first, smaller epidemic. This prevented case scenarios with 1/2 of their initial rat population being infected from completing more than 5 generations and scenarios with $I_r(0) = N_r(0)/100$ from completing more than 9 of the 11 generations.

We also completed a fit where we let $N_r(0) = 1$ and included the number of initial infected rats as a parameter to be fitted. Thus the zoonotic reservoir was represented in proportions of the population. The resulting fit (S1 appendix Fig 3 and 4) was still able to capture the seasonality of the data for Nigeria and the initial number of rats was estimated to be a median of 9.80% of the population (6.79–12.97%, 90% CI).

Discussion

Nigeria has experienced substantial Lf epidemics in recent years. The drivers of these epidemics have not, however, been previously explored in detail using a mechanistic model. The role of the rat reservoir in the seasonality of Lf cases is therefore not well understood. To that end, we developed and fitted a mechanistic model to national data of Lf epidemics in Nigeria from 2018 to 2020. This included previously over-looked rodent population dynamics for which the possibility of seasonal reproduction was investigated. We found that the model qualitatively replicated the weekly confirmed case data, showing that the seasonal peak in cases can be attributed to the population

and epidemiological dynamics of the zoonotic rodent reservoir.

By fitting our model to case data from the outbreaks in Nigeria, we inferred that the recruitment rate of *Mastomys* rats in contact with at-risk Nigerians was highly seasonal, which lead to rapid and substantial shifts in the proportion of the reservoir that were susceptible to and infected with Lf. We found that there was a pulse increase of new susceptible rats in December of each year that quickly become infected by LASV-carrying rodents surviving the previous epidemic. These infected rats then spread the infection to human inhabitants in shared rural environments. This finding corroborates the results of Coetzee (1965) and Coetzee (1975) that showed a marked increase in *Mastomys* breeding during the rainy season [14, 15]. This drove the increase in observed cases in humans with over 95% transmission events being spillover events from the zoonotic reservoir. This proportion was larger than that seen in other studies, such as Iacono et al (2015), which may reflect the omission of within hospital interactions and the increased risk of transmission in hospitals without appropriate infection control precautions [16].

In our sensitivity analysis, we adjusted the size of the vector population and the initial proportion of infected rodents. The total size of the vector population could be factored into the rat-to-human transmission parameter by a proportionate rescaling (see S1 appendix Fig 2(a)-(c) and 4 mid right) and the initial proportion of infected rats altered the simulated number of cases in a predictable manner. Altering initial values for infected rats predictably changed the trajectory of the simulated number of observed cases for the first year (see S1 appendix Fig 1 and 3) however simulation of data for subsequent years was unaffected.

Our estimates for the basic reproduction number—the expected number of secondary infections from a primary infection in a fully susceptible population—of Lf in rats were well above that which one would consider to be an extremely infective disease, e.g. measles being an order of magnitude lower [34]. This corresponded to the sharp increase in rat reproduction rates in December, when Nigeria typically experiences its dry season. This is in conflict with other studies that have demonstrated a strong correlation between *Mastomys* reproduction and rainfall with a delay of 1–3 months after the rainy season [12]. It is important to note that the model's representation of rat dynamics, particularly the recruitment function, may not capture the true complexity of the population dynamics. Using a recruitment function instead of a true birth function allows for a simplified life cycle of *Mastomys* by omitting a nesting/juvenile stage. Therefore dynamics that occur before recruitment and contact with humans may be missed, and could explain why the basic reproduction number for rats is higher than expected because potential susceptible individuals are excluded from transmission.

Incorporation of fine-scale space or further refinement of the rat population dynamics into the model—such as including time-dependent rat-to-rat transmission rates which may represent a hypothetical change in behaviour and proximity to humans—could improve both the model's realism and enhance our understanding of rat dynamics during different climate seasons. For example, these developments may explicitly consider increased rat populations near homes in the dry season and higher LASV prevalence in *Mastomys* during the rainy season [11, 35]. Additionally, a metapopulation model would better capture spatial variations between different administrative areas, improving the representation of disease dynamics. To enhance comprehensiveness, future studies should incorporate hospitalization and treatment options specific to human cases, offering insights into healthcare worker risks and transmission reduction strategies. These enhancements would advance understanding of the disease and improve applicability of the model, thereby helping guide better strategies for Lf control and prevention in Nigeria.

Conclusion

Our model captured the dynamics of weekly confirmed Lf cases over multiple epidemics in Nigeria. Our approach demonstrated that the population of *M. natalensis* experienced annual Lf outbreaks due to seasonal recruitment rates resulting in an influx of new susceptible rats to fuel the spread of the disease. It is not yet clear what proportion of this recruitment is due to seasonal reproduction or migration to rural homes in the dry season. The high number of infected vectors causes a spillover of infection, resulting in annual epidemics between late December and early April. There are only relatively low levels of human-to-human transmission, which supports the notion that the zoonotic reservoir of *M. natalensis* is the primary driver in the epidemiology of Lf. Therefore, we conjecture that the single most effective measure of controlling the epidemics would be to reduce human contact with the zoonotic reservoir, either by increased food security and hygiene, or with more effective trapping and culling of rats.

Supporting information

S1 Appendix. Appendix containing mathematical analysis and sensitivity analysis.

Acknowledgments

We thank the Nigeria Center for Disease Control for support and data supply.

References

1. World Health Organization. WHO Lassa fever fact sheet no 179; 2000.
2. Fichet-Calvet E, Rogers DJ. Risk maps of Lassa fever in West Africa. PLoS neglected tropical diseases. 2009;3(3):e388.
3. Monath TP, Newhouse VF, Kemp GE, Setzer HW, Cacciapuoti A. Lassa virus isolation from Mastomys natalensis rodents during an epidemic in Sierra Leone. Science. 1974;185(4147):263–265.
4. Demartini J, Green D, Monath T. Lassa virus infection in Mastomys natalensis in Sierra Leone: Gross and microscopic findings in infected and uninfected animals. Bulletin of the World Health Organization. 1975;52(4-6):651.
5. Ter Meulen J, Lukashevich I, Sidibe K, Inapogui A, Marx M, Dorlemann A, et al. Hunting of peridomestic rodents and consumption of their meat as possible risk factors for rodent-to-human transmission of Lassa virus in the Republic of Guinea. American Journal of Tropical Medicine and Hygiene. 1996;55:661–666.
6. McCormick JB, Webb PA, Krebs JW, Johnson KM, Smith ES. A prospective study of the epidemiology and ecology of Lassa fever. Journal of Infectious Diseases. 1987;155(3):437–444.
7. Ogbu O, Ajuluchukwu E, Uneke C, et al. Lassa fever in West African sub-region: an overview. Journal of vector borne diseases. 2007;44(1):1.
8. for Disease Control C, (CDC P, et al. Imported Lassa fever–New Jersey, 2004. MMWR Morbidity and mortality weekly report. 2004;53(38):894–897.

9. Akhmetzhanov AR, Asai Y, Nishiura H. Quantifying the seasonal drivers of transmission for Lassa fever in Nigeria. Philosophical Transactions of the Royal Society B. 2019;374(1775):20180268.

10. Zhao S, Musa SS, Fu H, He D, Qin J. Large-scale Lassa fever outbreaks in Nigeria: quantifying the association between disease reproduction number and local rainfall. Epidemiology & Infection. 2020;148:e4.

11. Fichet-Calvet E, Lecompte E, Koivogui L, Soropogui B, Doré A, Kourouma F, et al. Fluctuation of abundance and Lassa virus prevalence in Mastomys natalensis in Guinea, West Africa. Vector-Borne and Zoonotic Diseases. 2007;7(2):119–128.

12. Leirs H, Verhagen R, Verheyen W. The basis of reproductive seasonally in Mastomys rats (Rodentia: Muridae) in Tanzania. Journal of Tropical Ecology. 1994;10(1):55–66.

13. Leirs H, Verhagen R, Verheyen W, Mwanjabe P, Mbise T. Forecasting rodent outbreaks in Africa: an ecological basis for Mastomys control in Tanzania. Journal of Applied Ecology. 1996; p. 937–943.

14. Coetzee C. The breeding season of the multimammate mouse Praomys (Mastomys) natalensis (A. Smith) in the Transvaal Highveld. African Zoology. 1965;1(1):29–39.

15. Coetzee CG. The biology, behaviour, and ecology of Mastomys natalensis in southern Africa. Bulletin of the World Health Organization. 1975;52(4-6):637.

16. Lo Iacono G, Cunningham AA, Fichet-Calvet E, Garry RF, Grant DS, Khan SH, et al. Using modelling to disentangle the relative contributions of zoonotic and anthroponotic transmission: the case of Lassa fever. PLoS neglected tropical diseases. 2015;9(1):e3398.

17. Tobin EA, Asogun D, Akpede N, Adomeh D, Odia I, Gunther S, et al. Lassa fever in Nigeria: Insights into seroprevalence and risk factors in rural Edo State: A pilot study. Journal of Medicine in the Tropics. 2015;17(2):51.

18. World Health Organization. WHO list of blueprint priority diseases. 2022;.

19. Ibrahim MA, Dénes A. A mathematical model for Lassa fever transmission dynamics in a seasonal environment with a view to the 2017–20 epidemic in Nigeria. Nonlinear Analysis: Real World Applications. 2021;60:103310.

20. Musa SS, Zhao S, Gao D, Lin Q, Chowell G, He D. Mechanistic modelling of the large-scale Lassa fever epidemics in Nigeria from 2016 to 2019. Journal of theoretical biology. 2020;493:110209.

21. Bakare EA, Are EB, Abolarin O, Osanyinlusi SA, Ngwu B, Ubaka ON. Mathematical modelling and analysis of transmission dynamics of Lassa fever. Journal of Applied Mathematics. 2020;2020:1–18.

22. for Disease Control NC. An update of Lassa fever outbreak in Nigeria: situation reports; 2019.

23. Peel AJ, Pulliam J, Luis A, Plowright R, O'Shea T, Hayman D, et al. The effect of seasonal birth pulses on pathogen persistence in wild mammal populations. Proceedings of the Royal Society B: Biological Sciences. 2014;281(1786):20132962.

24. Diekmann O, Heesterbeek JAP, Metz JA. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. Journal of mathematical biology. 1990;28:365–382.

25. Van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical biosciences. 2002;180(1-2):29–48.

26. Group TWB. Available at https://data.worldbank.org/country/ng;.

27. Althaus CL. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLoS currents. 2014;6.

28. Sengupta P. The laboratory rat: relating its age with human's. International journal of preventive medicine. 2013;4(6):624.

29. Fichet-Calvet E, Becker-Ziaja B, Koivogui L, Günther S. Lassa serology in natural populations of rodents and horizontal transmission. Vector-Borne and Zoonotic Diseases. 2014;14(9):665–674.

30. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. Chapman and Hall/CRC; 1995.

31. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface. 2009;6(31):187–202.

32. Minter A, Retkute R. Approximate Bayesian Computation for infectious disease modelling. Epidemics. 2019;29:100368.

33. Beaumont MA, Cornuet JM, Marin JM, Robert CP. Adaptive approximate Bayesian computation. Biometrika. 2009;96(4):983–990.

34. Guerra FM, Bolotin S, Lim G, Heffernan J, Deeks SL, Li Y, et al. The basic reproduction number (R0) of measles: a systematic review. The Lancet Infectious Diseases. 2017;17(12):e420–e428.

35. Fichet-Calvet E, Lecompte E, Koivogui L, Daffis S, Meulen JT. Reproductive characteristics of Mastomys natalensis and Lassa virus prevalence in Guinea, West Africa. Vector-Borne and Zoonotic Diseases. 2008;8(1):41–48.

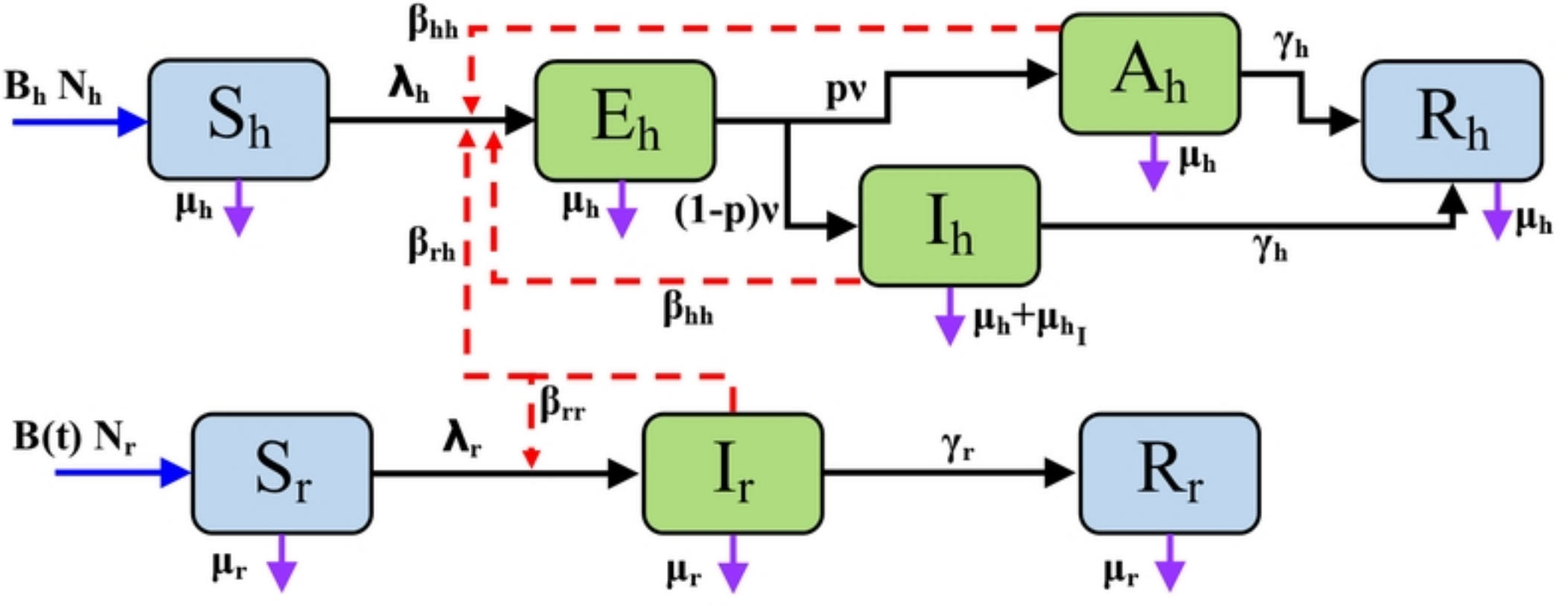

Figure 1, model schematic

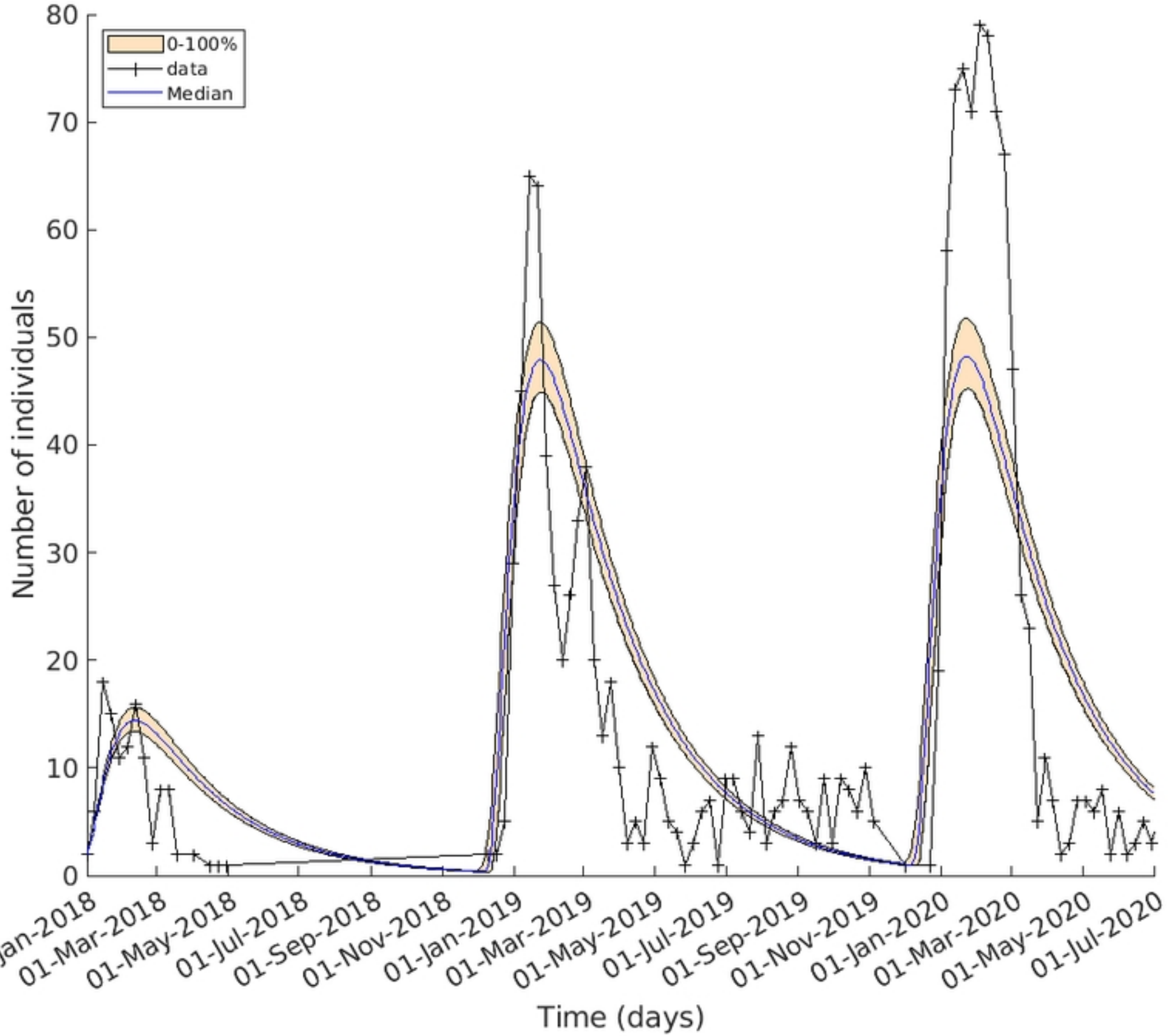

Figure 2, simulation

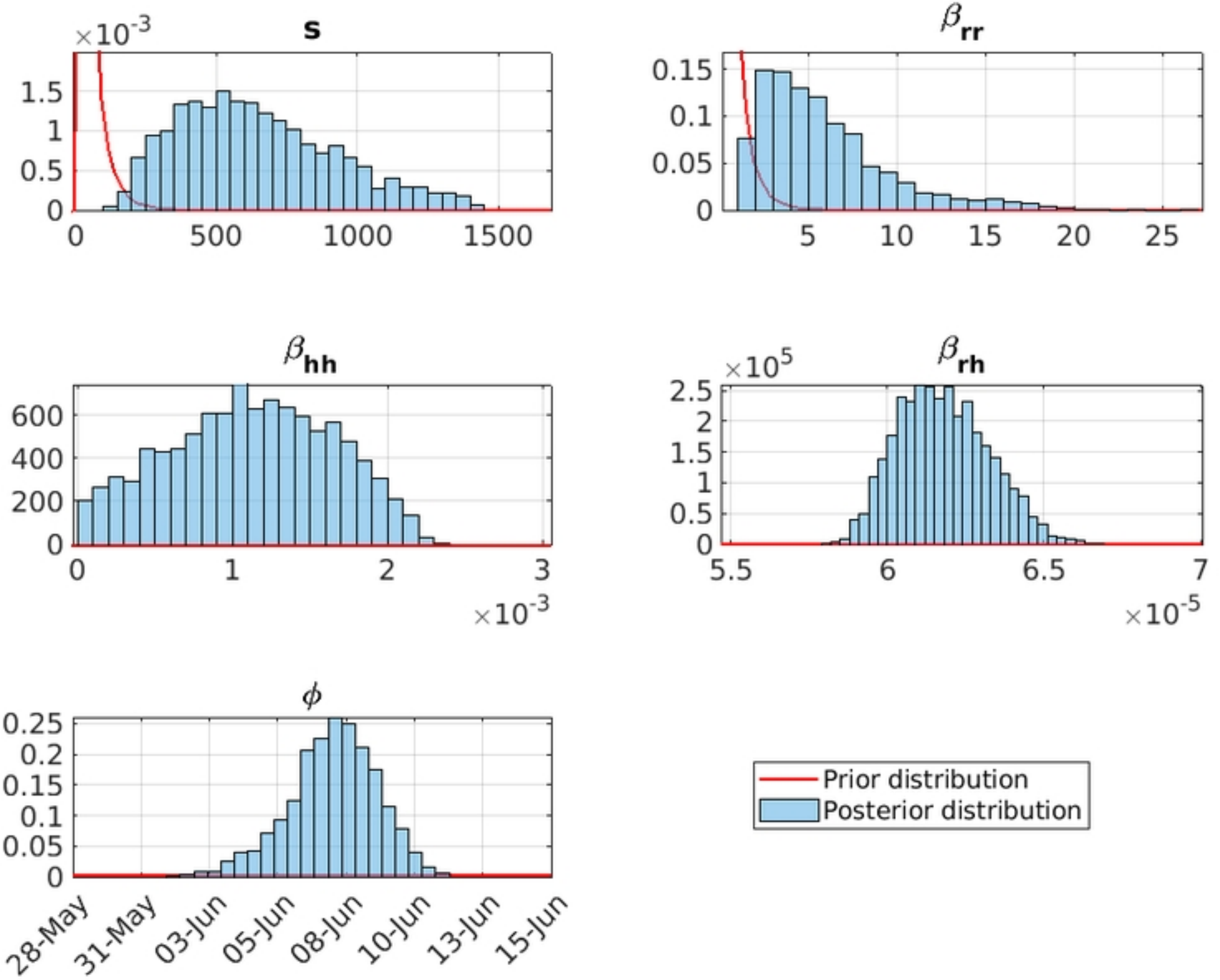

Figure 3, posterior distribution

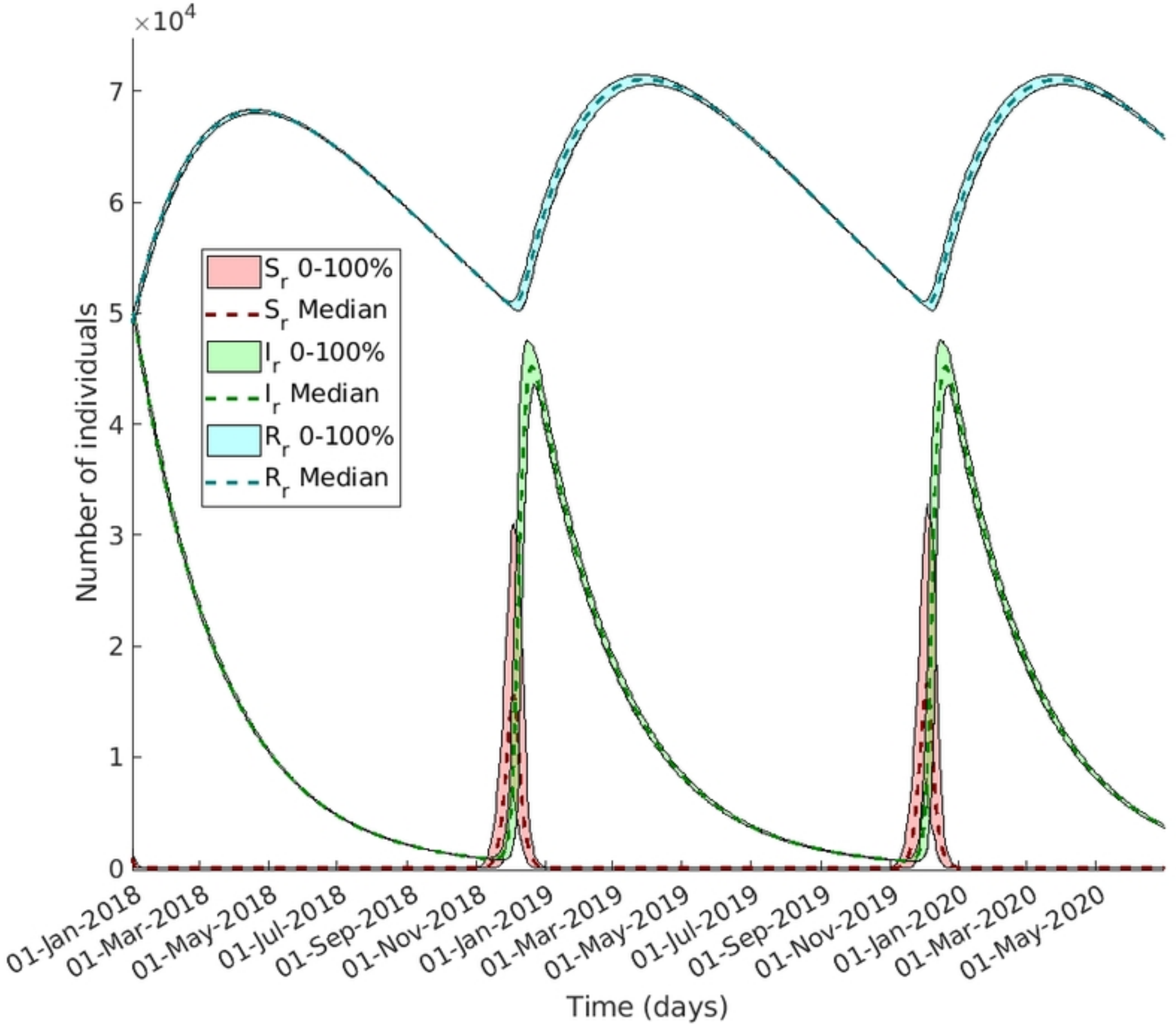

Figure 4, vector dynamics

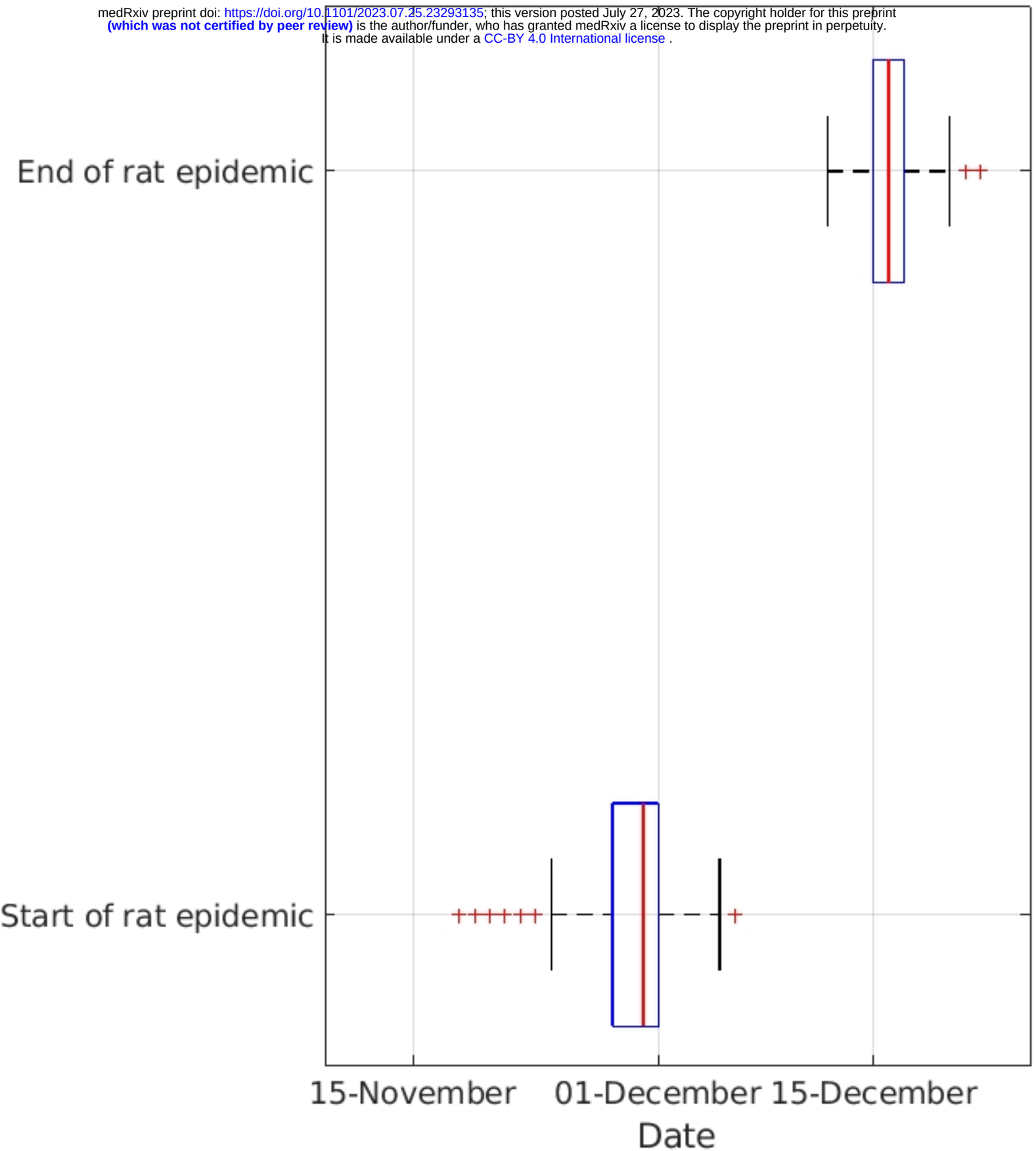

Figure 5, time-dependent R