MODELING INTRACELLULAR REPLICATION OF FLAVIVIRUSES

GABRIELA BLOHM

College of Public Health and Health Professions, University of Florida , Gainesville, FL

HAYRIYE GULBUDAK

Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA

MAIA MARTCHEVA

Department of Mathematics, University of Florida, Gainesville, FL

JOSE MIGUEL PONCIANO

Department of Biology, University of Florida, Gainesville, FL

Abstract. Recent technological advances have improved our knowledge of the replication process of Flaviviruses; however direct measurement of virus replication in the laboratory can be cost-prohibitive. Mathematical models of the within-host dynamics of virus replication can be useful for predicting virus load, which can affect virulence, transmission rate and rate of molecular evolution [1, 2]. In the case of arboviruses, most mathematical models have focused on transmission dynamics between hosts and vectors. Within-host replication of arboviruses (particularly Flaviviruses) has not been well-characterized mathematically. In this study, we model the replication of WNV in cell culture and generate testable predictions about the spread of the virus within a host. We study intracellular population dynamics of virus under a set of biologically sound parameter values and use statistical tools to quantify estimability of model parameters under distinct scenarios. The resulting mathematical model can also be applied in a clinical setting, where there is a growing body of work on the within-host dissemination of Flaviviruses.

Keywords: mathematical models, differential equations, flavivirus, West Nile Virus, Intracellular virus replication cycle, parameter estimation.

E-mail addresses: gblohm@ufl.edu, hayriye.gulbudak@louisiana.edu, maia@ufl.edu, josemi@ufl.edu.

Date: March 15, 2019.

1. Introduction

The genus Flavivirus of the family Flaviviridae contains more than 100 species of viruses. Clinically important members of this family include vector-borne viruses and others which are transmitted through direct contact. For example, Yellow Fever virus (YFV), Dengue virus (DENV), Japanese Encephalitis virus (JEV),West Nile virus (WNV) and Zika virus (ZIKV) are transmitted by mosquitos. Clinical symptoms of infection with these viruses include febrile illness and in some cases neurological disease [3]. West Nile virus (WNV) is the most widespread arbovirus in the world. It is transmitted by several species of mosquitos in North America [4] and has been detected in least 100 species of wild and domestic animals [5]. Due to widespread nature of Flaviviruses [6], it is important to understand the replication and transmission of these viruses in within-cellular, within-host and between-host scales. Here we focus on intracellular replication of Flaviviruses; in particular WNV.

Flaviviruses (Flaviviridae) are single-stranded RNA viruses that are approximately 12 kb in length, containing a capsid protein and a lipid envelope. At the cellular level, attachment and entry proceed by receptor-mediated endocytosis (primarily clathrin-mediated attachment), followed by uncoating in the cytoplasm, where replication and assembly take place. The virus particles then mature in the endoplasmic reticulum and are released as either infectious or defective virus particles [7]. Direct measurement of each step in the virus replication process is often cost-prohibitive. In cases where specific aspects of virus replication cannot be directly measured in the laboratory; therefore it is useful to work with mathematical models and statistical tools for parameter estimation.

Within-host replication of arboviruses (particularly Flaviviruses) has not been well-characterized mathematically. Several mathematical models of Hepatitis C virus (HCV: Flaviviridae) have captured several important aspects of virus replication, generating accurate predictions of within-host dissemination of the virus [8, 9] and of viral entry into host cells [10]. Kumberger et al. [9] model the within-host dynamics of HCV using a system of differential equations where the cells are distinguished according to their infection status (Uninfected, Infected, and Infectious), with viral entry, viral replication and viral export as rate parameters connecting the three state variables of the host cells. The model is used to predict cell-to-cell transmission in different types of tissue within a host. This approach proved effective at predicting the within-host spread of HCV that had been measured in the literature: in studies where viral maturation was measured by electron microscopy, and where clinical measurements of within-host dissemination of the virus were taken. In Padmanabhan et al. [10], the molecular process of viral entry was the focus of the model. The authors model the kinetics of virus entry into host cells by including CD81 expression as a determinant of viral entry. Cells resistant to infection due to reduced CD81 expression were included in the model, and as in the previous study, the mathematical model was able to predict results obtained in the literature.

In this study, we model the replication cycle of WNV following an approach that is similar to that of [9], with a few differences detailed below. We focus specifically on the virus population and assume a single state for the cells. We model the growth trajectory of a population of WNV particles in cell culture as a continuous process wherein virus

particles attach to the cell, replicate and mature inside the cell, and then are released into the supernatant as infectious virus particles. The rationale for this approach was to divide the replication cycle of the virus into distinct categories that could be measured in the laboratory, thus providing a model that can be tested and improved with published data on WNV replication dynamics and within-host proliferation. Furthermore, because these three processes represent common targets for antiviral therapies, mathematical models of the mechanisms that affect viral population size at each stage can inform the design of antiviral therapies.

This paper is organized as follows. In the next section, we present a coupled differential equation system describing the viral replication cycle depending on the location and infectivity of viral particles within a cell culture. In Sect. 3, we investigate the intracellular population dynamics of viral particles by exploring the equilibria and threshold analysis of the system. In Sect. 4, we use statistical tools to study estimability of model parameters under distinct scenarios. We point out knowing or fixing value of what parameter improves or gives the best estimates of the rest of the model parameters. In the last section, we summarize our results, and give the conclusion.

2. Model of Intracellular West Nile Virus Replication

We divide the virus population into three sub-populations (V, A, and R) according to their location and infectivity (See Fig.2). Upon contact with the host cell, infectious virus particles $V(t)$ bind to membrane receptors at time t and become incorporated via endocytosis at a rate α (stage A). The low pH of the endosome causes the virion envelope to fuse with the endosomal membrane, causing the nucleocapsid to become uncoated at a rate of σ virus particles per unit time. The viral RNA is then released into the cytoplasm, where RNA replication takes place at a rate ϕ. Virion assembly and maturation occur in the endoplasmic reticulum at a rate of β. Translation of viral proteins is the most energetically expensive process. Thus if the abundance of viral particles at stage A is large, due to over all consumed energy from the cell at this stage,and the large energy budget need for translation of viral proteins at stage R, we assume that the rate of assembly and maturation is weighted inversely by the concentration of attached virus particles: $\beta/(1+A(t))$ [11]. It is also well-known that WNV replication and maturation are error-prone processes, thus to account for the production of defective (non-infectious) virus particles, we include μ in the equation for $R(t)$ to denote the loss of virus particles from the system. Upon infectious virus particles entering the supernatant, either they become attached to a cell at a rate α or are degraded by RNases and other environmental factors at a rate δ. Figure 2 illustrates the trajectory of the virus population in each of the categories. In the schematic representation, all substages R_i, for $i = 1, 2, 3$, are represented by the stage R.

WNV establishes a persistent infection [12], meaning that a cell can be infected multiple times, producing several overlapping generations of virus particles before cell death.

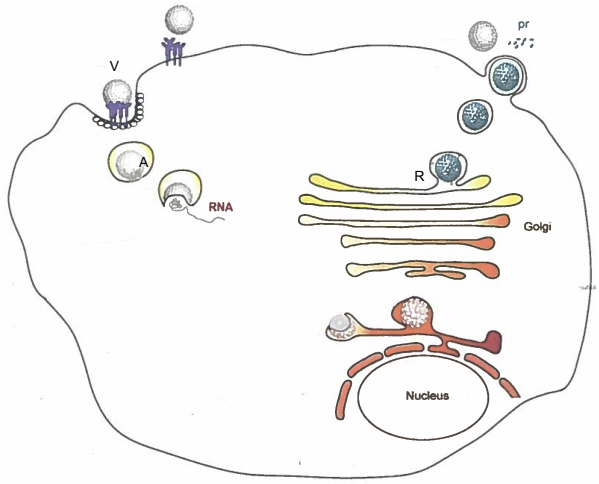

FIGURE 1. Schematic representation of a generalized Flavivirus replication cycle. Virus particles (V) become attached (A) to the cell membrane. Virus RNA is then released into the cytoplasm, where virus replication occurs. Newly assembled, immature virions enter the endoplasmic reticulum (R) and progress along a pH gradient until becoming released back into the extracellular matrix as infectious virus particles.

Hence we consider a continuous time model, consisting of a system of differential equations, describing the replication cycle of a virus with cell culture as follows:

$$\begin{cases} A'(t) = \alpha V(t) - \sigma A(t), \\ R'(t) = \phi\sigma A(t) - (\dfrac{\beta}{1+A(t)} + \mu)R(t), \\ V'(t) = \dfrac{\beta}{1+A(t)}R(t) - (\alpha+\delta)V(t) \end{cases} \tag{2.1}$$

2.1. **Equilibria and Stability Analysis of the Model.** Lets define the basic reproductive number for the system (2.1) as follows:

$$\mathcal{R}_0 = \phi\frac{\beta}{(\beta+\mu)}\frac{\alpha}{(\alpha+\delta)}. \tag{2.2}$$

The basic reproduction number is the average number of secondary viruses produced by one virus particle during its lifetime. The term $\dfrac{\alpha}{(\alpha+\delta)}$ in the $\mathcal{R}_0$ expression denotes the probability of free viruses attaching to a cell during its lifetime. The parameter ϕ is the average number of immature virus particles produced by one virus at stage R. Finally the term $\dfrac{\beta}{(\beta+\mu)}$ gives the probability of viral particle maturation in the absence of viruses at stage A.

The differential equation system given in (2.1) expresses the rate of change of the number of infectious virus particles per mL, the number of virus particles inside the cell as well as the number of attached virus particles. Setting the left hand side of the equations in the system equal to 0 and solving for the values of variables provide us

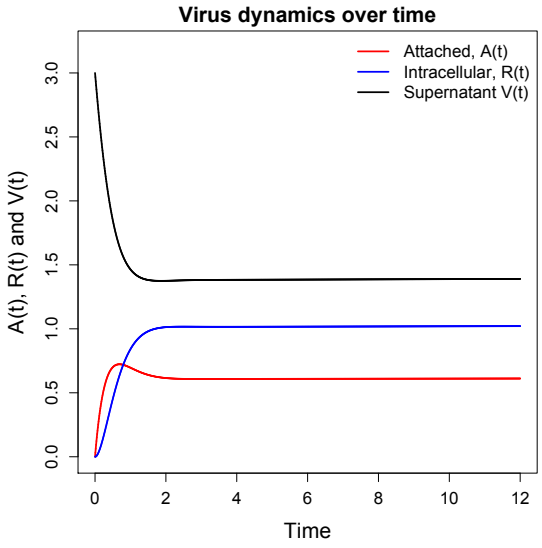

FIGURE 2. WNV population dynamics in cell culture. The number of virus particles per mL in the supernatant decreases exponentially and then stabilizes to a level that is lower than that of the original population size. The parameter values used here are as follows: $\alpha = 1.1,\ \sigma = 2.5, \phi = 1.00225,\ \beta = 1.5,\ \mu = 0.0001,\ \delta = 0.001$.

the values of the state variables for which their rate of change is 0. These are termed the equilibrium values of the state variables. The trivial equilibrium for the model is $\mathcal{E}^0 = (A^0 = 0, R^0 = 0, V^0 = 0)$. If however, all these equilibrium abundances are positive, such equilibrium is called an *endemic equilibrium*. In our case, the non-trivial equilibrium is $\mathcal{E}^* = (A^*, R^*, V^*)$, where

(2.3)
$$V^* = \frac{\phi\beta\alpha - (\beta+\mu)(\alpha+\delta)}{\frac{\alpha}{\sigma}\mu(\alpha+\delta)} = \frac{(\beta+\mu)\sigma(\mathcal{R}_0 - 1)}{\alpha\mu},\ A^* = \frac{\alpha}{\sigma}V^*,\ R^* = \frac{\phi\sigma A^*(1+A^*)}{\beta+\mu(1+A^*)}.$$

Note that whenever $\mathcal{R}_0 > 1$, the system has a positive endemic equilibrium. In biological setting, the equilibrium values can correspond to concentrations to which the system arrives some time after starting the experiment and subsequently stays there indefinitely, or they can correspond to concentration values that the system quickly can get away from despite starting with initial concentration values that are close to these equilibrium values of concentrations. In the first scenario, the model is said to have a *stable equilibrium*. In the second scenario, where the trajectories of the state variables depart from the equilibrium values, the model is said to admit an *unstable equilibrium*. The aim of our stability analysis is to determine under what conditions, these equilibria are stable.

Below, we show that the same condition: $\mathcal{R}_0 > 1$ must be met for the endemic equilibrium $\mathcal{E}^*$ to be stable. This stability condition is important because it delineates when a positive amount of free virus particles will persist in the system. In other words, this inequality represents an explicit condition that, if met, will guarantee the long-term

persistence of free virus particles in the system (as well as attached and intracellular).

Assuming that the initial concentration values (V_0, R_0, A_0) are sufficiently close to the equilibrium values in $\mathcal{E}^\dagger = (A^\dagger, R^\dagger, V^\dagger)$, with stability analysis, we aim to study the *long-term behavior of the deviations* of the system states from these equilibrium values. If the deviations grow indefinitely over time, the equilibrium is unstable. If however the deviations of the system state from the equilibrium values decay towards 0 over time, then the equilibrium is stable, i.e., the trajectories of the state variables $V(t), R(t), A(t)$ will eventually converge to their equilibrium values.

Let $x_1(t) = A(t) - A^\dagger$, $x_2(t) = R(t) - R^\dagger$ and $x_3(t) = V(t) - V^\dagger$ be all the time-dependent deviations of the trajectories from the equilibrium values. Let $\frac{dA(t)}{dt} = \mathcal{F}_1(A, R, V)$, $\frac{dR(t)}{dt} = \mathcal{F}_2(A, R, V)$, and $\frac{dV(t)}{dt} = \mathcal{F}_3(A, R, V)$. Then it follows that

$$\begin{cases} \frac{dx_1(t)}{dt} = \frac{d(A(t) - A^\dagger)}{dt} = \mathcal{F}_1(A, R, V), \\ \frac{dx_2(t)}{dt} = \frac{d(R(t) - R^\dagger)}{dt} = \mathcal{F}_2(A, R, V), \\ \frac{dx_3(t)}{dt} = \frac{d(V(t) - V^\dagger)}{dt} = \mathcal{F}_3(A, R, V) \end{cases} \tag{2.4}$$

Then, to study the behavior of the deviations from equilibrium over time, we approximate $\frac{dx_1(t)}{dt}$, for $i = 1, 2, 3$, by using a Taylor series expansion around the equilibrium $\mathcal{E}^\dagger = (A^\dagger, R^\dagger, V^\dagger)$ (in technical jargon we say that the system is *linearized*). Using vector notation by setting $x(t) = (x_1(t), x_2(t), x_3(t))$, the Taylor series approximation yields the following system of equations, written in matrix form:

$$\frac{dx}{dt} = \mathcal{J}(x) = \begin{pmatrix} \frac{\partial \mathcal{F}_1}{\partial A} & \frac{\partial \mathcal{F}_1}{\partial R} & \frac{\partial \mathcal{F}_1}{\partial V} \\ \frac{\partial \mathcal{F}_2}{\partial A} & \frac{\partial \mathcal{F}_2}{\partial R} & \frac{\partial \mathcal{F}_2}{\partial V} \\ \frac{\partial \mathcal{F}_3}{\partial A} & \frac{\partial \mathcal{F}_3}{\partial R} & \frac{\partial \mathcal{F}_3}{\partial V} \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}$$

Doing these partial derivatives gives

$$\frac{dx}{dt} = \mathcal{J}(x) = \begin{pmatrix} -\sigma & 0 & \alpha \\ \phi\sigma + \frac{\beta R^\dagger}{(1 + A^\dagger)^2} & -\left(\frac{\beta}{1 + A^\dagger} + \mu\right) & 0 \\ -\frac{R^\dagger \beta}{(1 + A^\dagger)^2} & \frac{\beta}{1 + A^\dagger} & -(\alpha + \delta) \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}$$

The sign of the eigenvalues λ_1, λ_2, and λ_3 of the Jacobian matrix $\mathcal{J}$ evaluated at the equilibrium determine the trajectories of those derivations because the solution trajectory for the deviations, written in vector format is of the form:

$$x(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} + c_3 e^{\lambda_3 t},$$

where the vectors c_i are constants depending on the model parameters [13]. When all the eigenvalues are negative, as t goes to ∞, all the terms $e^{\lambda_i t}$ converge to 0, and hence, all the elements in the vector of the deviations of equilibrium $x(t)$ converge to 0. In other words, as time goes to ∞, if all the eigenvalues of the Jacobian matrix are negative, then the system trajectories converge to the equilibrium values and the equilibrium is said to be stable. Knowing whether the three eigenvalues are negative amounts to verifying that the Routh-Hurwitz criterion [13] holds. For our three-dimensional model system, this criterion is stated as follow:
Let

$$\Delta = \left| \mathcal{J} - diag \begin{pmatrix} \lambda \\ \lambda \\ \lambda \end{pmatrix} \right| = \begin{vmatrix} -(\sigma+\lambda) & 0 & \alpha \\ \phi\sigma + \dfrac{\beta R^\dagger}{(1+A^\dagger)^2} & -\left(\dfrac{\beta}{1+A^\dagger} + \mu\right) - \lambda & 0 \\ -\dfrac{R^\dagger \beta}{(1+A^\dagger)^2} & \dfrac{\beta}{1+A^\dagger} & -(\alpha+\delta) - \lambda \end{vmatrix}$$

The determinant is readily found to be

$$\begin{aligned} \Delta &= -(\sigma+\lambda)\left(\frac{\beta}{1+A^\dagger} + \mu + \lambda\right)(\alpha+\delta+\lambda) \\ &\quad + \alpha\left[\left(\phi\sigma + \frac{\beta R^\dagger}{(1+A^\dagger)^2}\right)\frac{\beta}{1+A^\dagger} - \left(\frac{\beta}{1+A^\dagger} + \mu + \lambda\right)\frac{R^\dagger\beta}{(1+A^\dagger)^2}\right] \\ &= -(\sigma+\lambda)\left(\frac{\beta}{1+A^\dagger} + \mu + \lambda\right)(\alpha+\delta+\lambda) + \alpha\left[\phi\sigma\frac{\beta}{1+A^\dagger} - (\mu+\lambda)\frac{R^\dagger\beta}{(1+A^\dagger)^2}\right] \end{aligned}$$

Rearranging the expression above and setting it equal to zero, we obtain a third degree polynomial of λ: $a_0\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0$, where

$$\begin{aligned} a_0 &= 1, \\ a_1 &= \alpha + \sigma + \delta + \mu + \frac{\beta}{1+A^\dagger}, \\ a_2 &= (\frac{\beta}{1+A^\dagger} + \mu)(\sigma+\alpha+\delta) + \sigma(\alpha+\delta) + \alpha\frac{\beta R^\dagger}{(1+A^\dagger)^2} \\ a_3 &= \sigma(\frac{\beta}{1+A^\dagger} + \mu)(\alpha+\delta) - \alpha\phi\sigma\frac{\beta}{1+A^\dagger} + \mu\alpha\frac{\beta R^\dagger}{(1+A^\dagger)^2}. \end{aligned}$$

The zeros of the characteristic equation $\Delta = -a_0\lambda^3 - a_1\lambda^2 - a_2\lambda - a_3$ are the three eigenvalues of $\mathcal{J}$. The Routh-Hurwitz criterion states that these eigenvalues are all negative

(i.e., the equilibrium is stable) if $a_i > 0$, for all $i = 1, 2, 3$, and $a_1 a_2 - a_3 > 0$.

First equilibrium of interest is trivial equilibrium $\mathcal{E}^0$. Thus by setting $\mathcal{E}^\dagger = \mathcal{E}^0$, we obtain

$$\begin{aligned}
a_0 &= 1, \\
a_1 &= \alpha + \sigma + \delta + \mu + \beta, \\
a_2 &= (\beta + \mu)(\sigma + \alpha + \delta) + \sigma(\alpha + \delta) \\
a_3 &= \sigma(\beta + \mu)(\alpha + \delta) - \alpha\phi\sigma\beta = \sigma(\beta + \mu)(\alpha + \delta)[1 - \mathcal{R}_0]
\end{aligned}$$

It is clear that if $\mathcal{R}_0 < 1$, then $a_i > 0$, for all $i = 1, 2, 3$ and $a_1 a_2 - a_3 > 0$. Therefore by Routh-Hurwitz criterion, whenever $\mathcal{R}_0 < 1$, the trivial equilibrium $\mathcal{E}_0$ is (locally) asymptotically stable. Otherwise if $\mathcal{R}_0 > 1$, it is unstable since the system has at least one positive eigenvalue by Descartes' rule of sign.
Next equilibrium of interest is endemic equilibrium $\mathcal{E}^*$. Thus by setting $\mathcal{E}^\dagger = \mathcal{E}^*$, we obtain

$$\begin{aligned}
a_0 &= 1, \\
a_1 &= \alpha + \sigma + \delta + \mu + \frac{\beta}{1 + A^*}, \\
a_2 &= (\frac{\beta}{1 + A^*} + \mu)(\sigma + \alpha + \delta) + \sigma(\alpha + \delta) + \alpha\frac{\beta R^*}{(1 + A^*)^2} \\
a_3 &= \sigma(\frac{\beta}{1 + A^*} + \mu)(\alpha + \delta) - \alpha\phi\sigma\frac{\beta}{1 + A^*} + \mu\alpha\frac{\beta R^*}{(1 + A^*)^2} \\
&= \frac{\sigma}{1 + A^*}\left[(\beta + \mu(1 + A^*))(\alpha + \delta) - \alpha\phi\beta\right] + \mu\alpha\frac{\beta R^*}{(1 + A^*)} \\
&= \frac{\sigma}{1 + A^*}\left[(\beta + \mu)(\alpha + \delta) + \mu A^*(\alpha + \delta) - \alpha\phi\beta\right] + \mu\alpha\frac{\beta R^*}{(1 + A^*)} \\
&= \frac{\sigma}{1 + A^*}\left[\mu A^*(\alpha + \delta) + (\beta + \mu)(\alpha + \delta)[1 - \mathcal{R}_0]\right] + \mu\alpha\frac{\beta R^*}{(1 + A^*)} \\
&= \frac{\sigma}{1 + A^*}\left[\mu A^*(\alpha + \delta) - V^*\frac{\alpha\mu}{\sigma}(\alpha + \delta)\right] + \mu\alpha\frac{\beta R^*}{(1 + A^*)} \\
&= \mu\alpha\frac{\beta R^*}{(1 + A^*)}
\end{aligned}$$

It is clear that all coefficients a_0, a_1, a_2, a_3 are positive, whenever $\mathcal{R}_0 > 1$, since this condition implies the positivity of the equilibrium $\mathcal{E}^*$. Also note that $a_1 a_2 - a_3 > 0$. Then, by Routh-Hurwitz criterion, the endemic (positive) equilibrium $\mathcal{E}^*$ is locally asymptotically stable whenever the basic reproduction number $\mathcal{R}_0 > 1$, which completes the proof.

This model can be tested and improved with published data on WNV replication dynamics and within-host proliferation. Because these three processes in the replication

cycle represent common targets for antiviral therapies, the mechanisms, affecting viral population size at each stage can inform the design of antiviral therapies. Yet, obtaining values for all parameters in the laboratory poses a different challenge. However statistical tools and the results of the simulations can provide guidance on which parameters are most informative. In the next section, we investigate the estimability of the model parameters.

3. Assessing the estimability of the model parameters

We estimate a "true" value of each parameter and conduct simulations to determine the level of confidence in each parameter when varying sampling frequency. At the start of the experiment, the number of attached virus particles A_0 and the number of intracellular virus particles R_0 are both 0. We initiate virus population growth with a multiplicity of infection (MOI) of 3 virus particles per cell; thus $V_0 = 3$. To set the initial parameter values, we manually assigned a combination of values that would yield biologically feasible trajectories, where $V(t)$ would stabilize to a carrying capacity and $A(t)$ would remain below $V(t)$. The population in spent media ($V(t)$) is sampled at times $t_0, t_1, t_2, ..., t_n$ for a total duration T.

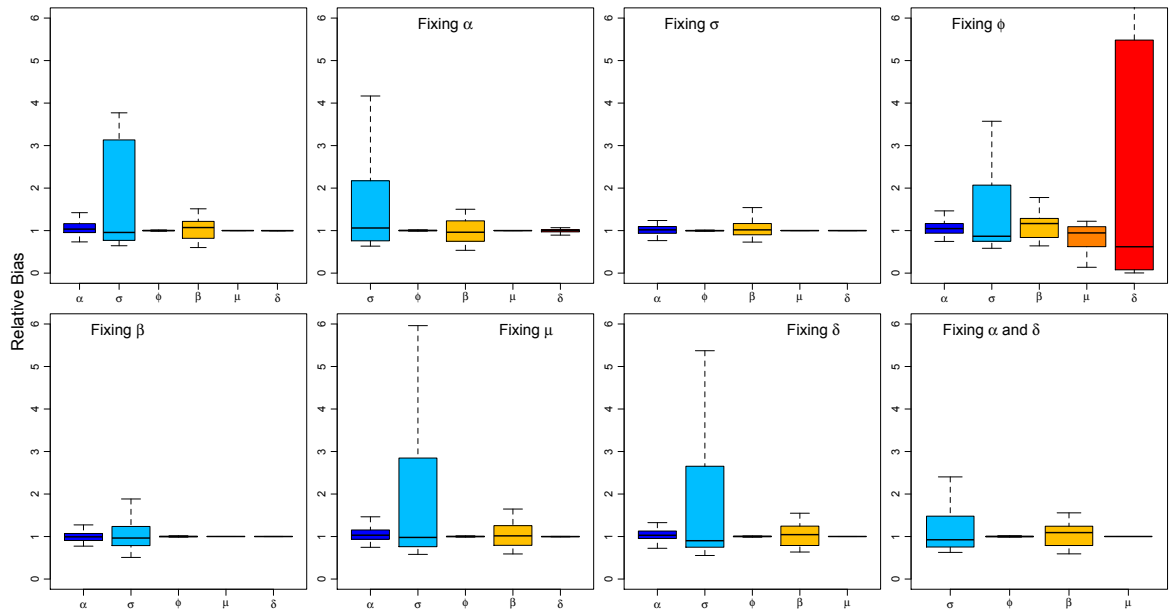

Figure 3. Relative bias of each parameter when certain parameters are known, after running 100 simulations for each set of parameters. In the upper left panel, it is assumed that all parameters are unknown. The most difficult parameter to estimate (most uncertain) in this model is β. Fixing (or "knowing") σ yields the lowest total amount of bias. When ϕ is fixed, the amount of uncertainty in the model increases.

3.1. **Materials and Methods.** The estimability of the model parameters is assessed in the context of the data generated by the serial passage experiment. Under this experimental setting, a typical data set consists of a number, say n, of virus samples collected

from the supernatant at regular time intervals (e.g. if samples are collected every 8 hours for a period of 72 hours that gives us 10 samples if time 0 is included). Then, to assess estimability of the model parameters, the idea is to simulate the trajectories of A, R, and V as well as regular random samples from those trajectories at intervals matching the experimental conditions. With those simulated samples, one can then estimate the model parameters and later judge, from the statistical properties of the estimates, the quality of the estimation. Furthermore, here we modified the sampling frequency in order to assess how does the quality of the estimate changes as a function of the number of samples taken, during the same total time period an experiment is running. The specific details of the simulations are given below.

The time series of samples with observation error were then used to estimate via least squares the model parameters. As stated in [14], least-squares estimation amounts to specifying a Normal likelihood. These two steps (simulation and estimation) are the basic building blocks of our simulation experiments. The statistical qualities of the estimates under any given experimental setting can be assessed by running these two steps a large number of times (here we used 100 times for all of our simulation settings). Then, the mean and variance of the estimates, relative to the true parameter values gives precise estimates of the bias and variability of the model parameter estimates.

We assessed the quality of the parameter estimates under two different simulation settings. In the first setting (simulated 100 times), we assumed that for a total of $T = 72$ hours, virus particles were sampled every 1.44 hours, to get a total of 50 observations. Although gathering 50 samples of the supernatant in 72 hours represents substantial experimental work, having a large number of samples gave us a benchmark for assessing the estimability of the parameters. Indeed, in time series statistics, the more samples over time are gathered, the better the quality of the estimate of the dynamics and of the model parameters are. The second experiment consisted of simulating a total of 25 observations taken every 2.88 hours during 72 hours. This setting corresponded to a realistic experimental setting. For this experimental setting, we also ran 100 simulation and estimation steps. Finally, for each sampling setting (25 and 50 samples), we assessed the quality of the estimates when one of the parameters was assumed to be known (or estimated empirically from a different experiment), for each one of the model parameters. Thus, for each one of the two settings we initially programmed 7 simulation experiments with 100 runs each (for a total of 700 simulation and estimation steps). The programs were all written in the language R and as it stands, the computer code takes about 24 hours to run per simulated setting.

3.2. **Numerical results.** The parameters that gave biologically feasible trajectories of the model are shown in Figure (2). The first subfigure in Figure (3) presents a simple test of estimability of the model parameters, using 50 simulations. To obtain this figure, we conducted 50 simulations of the A, R, V trajectories with 25 where the deterministic trajectories were contaminated each time with different levels of observation error, (up to 20%). For each simulated time series, we estimate the model parameters using only as observations the time series of V contaminated with observation error. The figure shows that all the estimated parameters fall around the true value, which gave us a quick indication that the model parameters were indeed estimable when one has as data

only the time series of the free virus particles. Going from 25 to 50 samples in time only improves the quality of the inference by a little bit, but not substantially (results not shown). When estimating all parameters, all appear unbiased. The variance of the parameter estimates, however, varies widely which indicates that the time series of virus particles contains more information for some parameters than others. For example, Figure (3) shows that σ has a large variance and is sometimes over-estimated three to fourfold. Because σ is the most difficult parameter to estimate, when this parameter is known, the remaining parameters are estimated with the highest precision and unbiased for all experiments. Assuming that ϕ is known, it yields the worst estimates. Fixing β also gives good parameter estimates, but β would be difficult to measure empirically. Fixing μ, δ, and α individually improves the precision the estimates of σ and more so when α is the parameter that is fixed.

4. Conclusion

Laboratory experiments and mathematical models can facilitate progress towards controlling the spread of arboviruses; however in the absence of field research, the results from these studies cannot be ground-tested and the models cannot be improved. Conducting clinical research in tandem with the development of these models is important. Here we introduced a coupled differential equation system describing three stages of virus replication cycle depending on the location and infectivity of virus particles. The model analysis conveys important qualitative predictions. First, we derive a biologically interpretable basic reproductive number $\mathcal{R}_0$ for the intra-cellular dynamics of the WNV. It is an important threshold value informing how the persistence of the infectious virus particles can be affected when one or several of these processes is changed. It provides quantitative predictions such as under what conditions, virus persistence occurs or virus population dies out. Next, we show that whenever $\mathcal{R}_0 < 1$, the virus dies out provided that initial free virus population is sufficiently small ($\mathcal{E}^0$ is locally asymptotically stable). Otherwise if $\mathcal{R}_0 > 1$, the viruses persists provided that initial abundance A_0, R_0, V_0 are sufficiently close to the endemic equilibrium $\mathcal{E}^* = (A^*, R^*, V^*)$ ($\mathcal{E}^*$ is locally asymptotically stable). These predictions can be scaled up in order to link the intra cellular processes with processes at higher scales involving the vector itself.

Estimation of the parameters of a biological, dynamic model using time series data is a common approach in ecological research [15, 16, 17]. We apply this approach in the field of virology, where we attempt to estimate parameters in the virus replication cycle. We show that when the data at hand consists of time series of observations of one of the stages of the virus replication cycle ($V(t)$), it is possible to estimate with a suitable degree of precision the value of the critical parameters governing the transitions and changes in the virus life-history processes. Although the model is a first and simple representation of the WNV life-cycle, our results represent a substantial step into the formulation of a general understanding of the dynamical processes involved in such life-cycle.

In general, all parameter estimates are unbiased; however the precision varies according to which parameter is known (Figure (3)). When fixing ϕ, there is a large reduction

in the precision of the model. This parameter appears only once in the system of equations and is multiplied by σ. Thus, fixing it imposes a numerical restriction on the potential optimal solution. These results emphasize that teasing apart the estimates of ϕ and σ can be challenging. If one could estimate α and δ, one can reduce the variability in the estimates of σ so that now the estimates are on average unbiased and with a precision that frames the estimate within 0.5 and 1.5× relative to the true value (Figure (3)).

Obtaining values for the parameters in the laboratory poses a different challenge; however the results of the simulations provide guidance on which parameters are most informative, all else being equal. In an experimental setting, $V(t)$ can be measured directly as plaque-forming units or, alternatively, by TCID50. To determine the number of defective virus particles that are being produced, the difference between the number of genome copies by real-time RT-PCR and the number of plaque-forming units can be calculated. This will allow an estimation of μ, the rate at which virus particles leave the system either via incomplete maturation or assembly, or due to degradation of viral RNA. Measuring $R(t)$ becomes more costly: electron microscopy can be used to visualize the process of virus particle maturation through the endoplasmic reticulum. Direct measurement of $A(t)$ and of the kinetics of RNA replication in the cytoplasm (particularly σ) is challenging. By empirically measuring the rate of virus attachment and maturation, it will be possible to obtain statistical estimates of the parameters that cannot be measured in the laboratory.

Our simulations and assessment of the statistical properties of the Maximum Likelihood (ML) estimates of the model parameters show *how to target efforts aiming at complementing time-series statistical analyses with experimental methods.* Because the variance of the parameters varies widely from one parameter to the other, the changes in the observed time series of free virus particles contains more information to estimate some parameters than others. Via our approach where we assumed that one of the parameters at a time was known, we were able to isolate the set of parameters that it would be more valuable to estimate using experimental work. We hope that the results presented here offer guidance and ideas as to which one of these parameters or intra-cellular life-cycle stages, it would be most interesting to target in the laboratory.

References

[1] Kotzin, J. J., Mowel, W. K., Henao-Mejia, J. (2017). Viruses hijack a host lncRNA to replicate. *Science*, **358(6366)**, p. 993–994.

[2] Wang, P., Xu, J., Wang, Y., Cao, X. (2017). An interferon-independent lncRNA promotes viral replication by modulating cellular metabolism. *Science*, **358(6366)**, p.1051–1055.

[3] Petersen, L. R., Marfin, A. A. (2005).Shifting epidemiology of Flaviviridae. *Journal of travel medicine*, **12(s1)**.

[4] Turell M. J., Dohm D. J., Sardelis M. R., Oguinn M. L., Andreadis T. G., Blow J. A. (2005). An Update on the potential of North American Mosquitoes (Diptera: Culicidae) to Transmit West Nile Virus. *J Med Entomol.*, **42(1)**, p.57–62.

[5] Root, J. J. (2013). West Nile Virus associations in wild mammals: a synthesis. *Archives of Virology*, **158**, p.735–752.

[6] Crowder D. W., Dykstra E. A., Brauner J. M., Duffy A., Reed C., et al. (2013). West Nile Virus Prevalence across Landscapes Is Mediated by Local Effects of Agriculture on Vector and Host Communities. *PLoS ONE*, **8(1)**.

[7] Fernandez-Garcia, M. D., Mazzon, M., Jacobs, M., Amara, A. (2009). Pathogenesis of flavivirus infections: using and abusing the host cell. *Cell host & microbe*, **5(4)**, p.318–328.

[8] Dahari H., Ribeiro, R. M., Rice, C.M., Perelson, A.S. (2016). Mathematical Modeling of Subgenomic Hepatitis C Virus Replication in Huh-7 Cells. *Journal of Virology*, **81(2)**, p.750–760.

[9] Kumberger, P., Frey, F., Schwarz, U. S. and Graw, F. (2016), Multiscale modeling of virus replication and spread. *FEBS letters*, **590(13)**, p.972–1986.

[10] Padmanabhan, P., Dixit, N. M. (2011). Mathematical Model of Viral Kinetics In Vitro Estimates the Number of E2–CD81 Complexes Necessary for Hepatitis C Virus Entry. *PLoS Computational Biology*, **7(12)**,p. e1002307. https://doi.org/10.1371/journal.pcbi.1002307.

[11] Mahmoudabadi, G., Milo, R., Phillips, R. (2017). Energetic cost of building a virus. *Proceedings of the National Academy of Sciences*, p.201701670.

[12] Franz, A. W. E., Kantor, A. M., Passarelli, A. L., Clem, R. J. (2015). Tissue barriers to arbovirus infection in mosquitoes. *Viruses*, **7(7)**, p. 3741–3767.

[13] Kot, M. (2011). *Elements of mathematical ecology. Cambridge University Press.*

[14] Ponciano, J. M., Capistrán, M. A. (2011). First principles modeling of nonlinear incidence rates in seasonal epidemics. *PLoS computational biology*, **7(2)**, p.e1001079.

[15] Ferguson, J.M., Ponciano J.M. (2014). Predicting the process of extinction in experimental microcosms and accounting for interspecific interactions in single-species time series. *Ecology Letters*, **17**, p.251-259.

[16] Ferguson, J., Carvalho, F., Murillo, O., Ponciano J.M. (2015). Population growth, density dependence and evironmental autocorrelation in animal populations. *Theoretical Ecology*, **8(2))**, doi:10.1007/s12080-015-0276-6.

[17] Ferguson, J.M., Ponciano J.M. (2015). Evidence and implications of higher order scaling in the environmental variation of animal population growth. *Proceedings of the National Academy of Sciences*, **112(9)**, doi: 10.1073/pnas.1416538112.