Calibration of a SEIR-SEI epidemic model to describe the Zika virus outbreak in Brazil
Eber Dantas, Michel Tosin, Americo Cunha Jr

To cite this version:

HAL Id: hal-01456776
https://hal.archives-ouvertes.fr/hal-01456776v3
Submitted on 27 Sep 2017 (v3), last revised 4 Jul 2018 (v4)
Calibration of a SEIR–SEI epidemic model to describe the Zika virus outbreak in Brazil

Eber Dantas\(^a\), Michel Tosin\(^a\), Americo Cunha Jr\(^a,*\)

\(^a\)NUMERICO – Nucleus of Modeling and Experimentation with Computers, Universidade do Estado do Rio de Janeiro (UERJ), Rua São Francisco Xavier, 524, Rio de Janeiro - RJ, 20550-900, Brasil. Phone: +55 21 2334-0323 r: 208

Abstract

Multiple instances of Zika virus epidemic have been reported around the world in the last two decades, turning the related illness into an international concern. In this context the use of mathematical models for epidemics is of great importance, since they are useful tools to predict the underlying outbreak numbers and allow one to test the effectiveness of different strategies used to combat the associated diseases. This work deals with the development and calibration of an epidemic model to describe the 2016 outbreak of Zika virus in Brazil. A system of 8 differential equations with 7 parameters is employed to model the evolution of the infection through two populations. Nominal values for the model parameters are estimated from the literature. An inverse problem is formulated and solved by comparing the system response to real data from the outbreak. The calibrated results presents realistic parameters and returns reasonable predictions, with the curve shape similar to the outbreak evolution and peak value close to the highest number of infected people during 2016. Considerations about the lack of data for some initial conditions are also made through an analysis over the response behavior according to their change in value.

Keywords: Zika virus dynamics; nonlinear dynamics; mathematical biology; SEIR epidemic model; model calibration; system identification

*Corresponding author.

Email addresses: eber.paiwa@uerj.br (Eber Dantas), michel.tosin@uerj.br (Michel Tosin), americoc@ime.uerj.br (Americo Cunha Jr)
1. Introduction

The Zika virus is a flavivirus that upon infection in humans causes an illness, known as Zika fever, identified commonly with macular or papular rash, mild fever and arthritis [1, 2]. It is mainly a vector-borne disease carried by the genus *Aedes* of mosquitoes [2, 3], while in a lesser amount it is also transmitted via sexual interaction [4, 5], and contamination by blood transfusion is under investigation [6]. The Zika virus was first isolated in primates from the Zika forest in Uganda in 1947 [7]. Evidences of the virus in humans were found in Nigeria in 1968 [8]. An epidemic occurred in 2007 on Micronesia [9], followed by multiple outbreaks on several Pacific Islands between 2013 and 2014 [10, 11]. The first Zika virus autochthonous case in Brazil was reported around April, 2015 [12], and nearly 30,000 cases of infection were already notified by January 30, 2016 [13], along with the Pan American Health Organization being informed in the same month about locally-transmitted cases on numerous continental and island territories of America [14]. The Brazilian Ministry of Health registered 215,319 probable cases of Zika fever (130,701 of which were confirmed) until the 52th epidemiological week (EW) of 2016 [12]. The Zika epidemic has been causing concern in the international medical community, health authorities and population, specially due to an association between the Zika virus and other diseases, such as newborn microcephaly [4, 15] and Guillain-Barré syndrome [16], whose correlation to the Zika virus was considered by the World Health Organization a “scientific consensus” [17].

In this epidemic scenario, the development of control and prevention strategies for the mass infection is a critical issue. A mathematical model able to predict the number of infected people during the virus outbreak is a useful tool, which can be employed to identify effective and vulnerable aspects on disease control programs [18, 19, 20]. Furthermore, for an epidemic model to be truly useful it must undergo a judicious process of validation [21, 22], which consists in comparing model predictions with real data in order to evaluate if they are realistic. In general, the first predictions of a model do not agree with the observations, possibly due to inadequacies in the model hypotheses or because of a poor choice for the model nominal parameters. The first case invalidates the model, but the second can be amended through a procedure known as model calibration, where a set of parameters that promote a good agreement between predictions and observations is sought.

This work is one of the results in a rigorous ongoing process of identification and validation of representative models to describe Zika virus outbreaks in Brazil [23, 24]. For this purpose, a SEIR-SEI mathematical model is adapted to the Brazilian scenario based on similar SEIR descriptions used successfully for other Zika outbreaks that happened around the world [25, 26], and on studies performed over SEIR dynamical systems [27, 28, 29, 30]. The nominal values of the model parameters belong to characteristics of the Zika infection and its vector, quantitatively estimated in the literature or published by health organizations. Predictions are obtained from numerical simulation and further heuristic manipulation, followed by a comparison to real data of the outbreak as an initial
effort to validate the model. In sequence, a rigorous process of model calibration is performed through the formulation and solution of an inverse problem.

The rest of this paper is organized as follows. In section 2, the mathematical model is described and estimation of nominal values for the model parameters is discussed. In section 3, the forward and inverse problems are formulated and solved, where results are detailed and a subsequent comparison between model predictions and experimental data is made to calibrate the model. Finally, in section 4, the main contributions of this work are emphasized and a path for future works is suggested.

2. Epidemic model for Zika virus dynamics

2.1. Model hypotheses

This work utilizes a variant of the Ross-Macdonald model [31] for epidemic predictions, separating the populations into a SEIR-SEI framework (susceptible, exposed, infectious, recovered) [32, 33, 34]. Each category represents the health condition of an individual inside such group at time \( t \), with respect to the considered infection. The susceptible group, denoted by \( S(t) \), represents those who are uncontaminated and are able to become infected. The exposed portion of the population, \( E(t) \), comprehends anyone that is carrying the pathogen but is still incapable of transmitting the disease. While the infectious individuals, \( I(t) \), can spread the pathogen and may display symptoms associated with the illness. Finally, the recovered group, \( R(t) \), contains those who are no longer infected.

The populations of humans and vectors are segmented into the aforesaid classes (excepting the group of recovered vectors), as Figure 1 depicts schematically with the accordingly subscripts. \( S_h, E_h, I_h \) and \( R_h \) amass the number of people at each stage of the model description, and \( S_v, E_v, I_v \) signifies proportion of vectors (0 \( \leq \) \( S_v, E_v, I_v \leq 1 \)).

Demographical changes in the number of humans are not considered and the total vector population is maintained constant during the analysis, although
variations of the proportion of vectors on the particular compartments are introduced via birth and death rates. The vector in question is regarded as a hypothetical mosquito apt to being infected or infectious throughout all its lifetime — which means the model accounts only for the adult stage of their life cycle — and also unable to recover.

The time elapsed while an individual is on the aforementioned exposed group is known as the latent period of an organism and, in this work, is adopted as equivalent to the commonly called incubation period (the time elapsed between being infected and exhibiting symptoms), since data is extremely sparse on the latter for humans [35] (namely, the intrinsic incubation period). Both terms are used interchangeably hereafter, and the concepts do not differ on the mosquitoes case (the extrinsic incubation period) [36]. In addition, all the members of the susceptible group are treated as equally capable of being infected and the recovered ones as completely immunized.

2.2. Model equations

The evolution of individuals through the SEIR-SEI groups is governed by the following (nonlinear) autonomous system of ordinary differential equations

\[
\begin{align*}
\frac{dS_h}{dt} &= -\beta_h S_h I_v, \\
\frac{dS_v}{dt} &= \delta - \beta_v S_v I_h N - \delta S_v, \\
\frac{dE_h}{dt} &= \beta_h S_h I_v - \alpha_h E_h, \\
\frac{dE_v}{dt} &= \beta_v S_v I_h N - (\alpha_v + \delta) E_v, \\
\frac{dI_h}{dt} &= \alpha_h E_h - \gamma I_h, \\
\frac{dI_v}{dt} &= \alpha_v E_v - \delta I_v, \\
\frac{dR_h}{dt} &= \gamma I_h, \\
\frac{dC}{dt} &= \alpha_h E_h,
\end{align*}
\]  

(1)

where \(N\) represents the total human population, \(1/\alpha\) is the disease’s incubation period (each with the corresponding subscript of \(h\) for human’s and \(v\) for vector’s), \(1/\beta\) means the vector lifespan, \(1/\gamma\) is the human infection period — which is defined in this work as the interval of time that a human is infectious — and \(\beta\) identifies the transmission rate, specifically \(\beta_h\) is the mosquito-to-human rate and \(\beta_v\) the human-to-mosquito rate. The first delta on the right hand side of the \(dS_v/dt\) equation sustains the hypothesis that the total vector population is unchanged during the analysis, and the \(dC/dt\) equation allows evaluation of the cumulative number \(C(t)\) of infectious people until the time \(t\), that is, the amount of humans so far that contracted the disease and have passed through or are in the infectious group at the given time.

Additionally, a set of 52 points to represent the number of new infectious cases of Zika fever at each week is defined as follows:

\[
N_w = C_w - C_{w-1}, \quad N_1 = C_1,
\]  

(2)

where \(C_w\) is the cumulative number of infectious humans in the \(w\)th EW.
2.3. Nominal parameters

The preliminary values for the parameters of the set of Eqs. (1) come from the related literature concerning the Zika infection, vector-borne epidemic models, the *Aedes aegypti* mosquito (which is the main vector for Zika, Dengue and Yellow fever in Brazil) and publications provided by health organizations and government agencies. The Brazilian Institute of Geography and Statistics reports that Brazil had approximately $N = 206 \times 10^6$ people by July, 2016 [37]. The adopted extrinsic incubation period is $1/\alpha_v = 9.1$ days [25]. This value agrees with other statistical confidence intervals (CI) that are presented for the parameter in another works (95% CI: 7.3–9.3 days [38]) and is close to the numbers suggested by experimental studies for the time necessary for the virus to reach the mosquito’s saliva after an infectious blood meal (5 [39, 40] and 7 days [41]). A systematic review and pooled analysis of the literature and case studies available in [42] estimates that the median intrinsic incubation period is 5.9 days (95% CI: 4.4—7.6). This values is selected for $1/\alpha_h$ in this work and is compatible with the range of 3–12 days recommended by multiple source [15, 43, 44], also formerly used in previous studies [45]. The aforementioned literature analysis in [42] also concludes that 9.9 (95% CI: 6.9–21.4) days is the mean time until an infected has no detectable virus in blood. Considering the assumption that the infectiousness in Zika infection ends $1.5–2$ days before the virus becomes undetectable [25, 38], the chosen value for the human infectious period is $1/\gamma = 9.9 – 2 = 7.9$ days. As for the vector lifespan $1/\delta$, “the adult stage of the mosquito is considered to last an average of eleven days in the urban environment” according to [46]. This is the assumed value for the parameter in this work, which is also consistent with the usual life expectancy for the mosquito in Rio de Janeiro, Brazil [47], and comes close to the average of 2–3 weeks considered in biological studies about the species [48] and by the Centers for Disease Control and Prevention [49]. Lastly, the time between a mosquito being infected and it infecting a human, $1/\beta_h$, and the time between a human infection and a mosquito taking an infectious blood meal, $1/\beta_v$, is estimated in [38] as an average of $11.3$ days (95% CI: 8.0–16.3) and $8.6$ days (95% CI: 6.2–11.6), respectively.

3. Calibration of the epidemic model

3.1. Forward Problem

The evaluation of the system response is the forward problem associated to the model presented in section 2 and is performed numerically in this work via a Runge-Kutta (4,5) method. Equation 2 provides a scalar observable of interest, used to assess the simulation when compared with real data of the 2016 outbreak made available by the Brazilian Ministry of Health [50] (supplementary material A). The referred data consists of probable cases of infected people per EW, registered by sanitary outposts and health institutions throughout the country when the common symptoms of Zika fever were exhibited by an individual. In accordance with the hypothesis that one only displays symptoms when inside
$I_h(t)$, the $C(t)$ variable models this discrete accumulating data on a continuous sense and $N_w$ provides the corresponding influx per EW. The $C(t)$ time series is also observed as a criteria for good fitting, since it signifies the general impact of the epidemic and because a reasonable $N_w$ result does not necessarily implies an acceptable cumulative number of infectious individuals for all $t$ in comparison to the data.

Numerical integration requires proper definition of the dynamical system underlying the SEIR-SEI epidemic model by means of prescribing the values of its variables at the initial time of the analysis, established as the first EW of 2016. The remaining individuals in both populations are assumed susceptible at first, meaning $S^i_h = N - E^i_h - I^i_h - R^i_h$ and $S^i_v = 1 - E^i_v - I^i_v$. The initial values for the exposed and infectious groups are assumed the same, $E^i_h = I^i_h$ and $E^i_v = I^i_v$. Likewise, the number of infected humans at the initial time must be $I^i_h = C^i$, given its definition. The value of $C^i$ is taken as the number of confirmed Zika cases in Brazil on the first EW of 2016 [50], 8,201 individuals, and the recovered group is assumed equal to the suspected number of infected in 2015, according to the data available [13], $R^i_h = 29,639$ individuals. As for the proportion of infectious vectors in the first week, to work around the lack of data for this initial condition (IC), repetitive manual estimations were tried until the resulted time series of $N_w$ presented reasonable values compared to the real data. It became clear that the system response is very sensible to $I^i_v$, as slight variations in its value are required to achieve feasible results. In the process of choosing its value, the matching of the $N_w$ curve’s peak to the amplitude of infection is also a priority, since this is the main interest region for evaluation of the outbreak. The nominal values of the parameters exhibited viable $N_w$ curves around $I^i_v = 2.2 \times 10^{-4}$.

Figure 2 presents the SEIR-SEI model response for the nominal set of parameters from section 2.3, supplied with the above IC, on an epidemiological week temporal domain consisting of one to fifty-two weeks (7 to 365 days), compared with the data of the outbreak (red dots).
The general shape of the curves in Figure 2 do provide qualitative information regarding the evolution of the infection, even though the predictions portrayed are not quantitatively realistic. This inherent pattern agreement and numerical mismatch suggest that the model response may differ from the real data due to the use of unsuitable values for the parameters or incorrect IC assumptions. The search for parameter and initial condition values that make the simulations fit well to the observed data defines model calibration (or system identification), being the object of interest of the next section.

3.2. Inverse Problem

The epidemic model of section 2 is a continuous-time dynamical system of the form

$$\dot{x}(t) = f(x(t), p), \quad x(t_0) = x_0 \quad (3)$$

where $x(t) \in \mathbb{R}^8$ is the vector of states at time $t$, $x_0 \in \mathbb{R}^8$ is a prescribed initial condition vector referring to the initial time $t_0$ of the analysis, the vector $p \in \mathbb{R}^7$ lumps the model parameters and $f: \mathbb{R}^8 \times \mathbb{R}^7 \rightarrow \mathbb{R}^8$ is a nonlinear map that defines the laws of evolution of this dynamics.

The forward problem exemplified on the previous section consists in providing initial conditions and a set of parameters, represented by the pair $\alpha = (x_0, p)$, and compute by means of numerical integration the model response $x(t)$ from which a scalar observable $\phi(\alpha, t)$ is obtained.

The model calibration problem seeks to find a set of parameters that, to a certain degree, makes the response of the model as close as possible to the empirical observations (reference data). In formal terms, given a data set with $M$ scalar observations $y_1, y_2, \ldots, y_M$, the calibration aims at finding a set $\alpha^*$ such that

$$\alpha^* = \arg \min_{\alpha} \left\{ \sum_{m=1}^{M} \left| y_m - \phi(\alpha, t_m) \right|^2 \right\}, \quad (4)$$

where $\{t_m\}_{m=1}^{M}$ corresponds to a discrete set of time-instants to which the $M$ scalar observations are assigned. This is the inverse problem associated to the epidemic model. In general this type of problem is extremely nonlinear, with none or low regularity, multiple solutions (or even none), being much more complicated to attack in comparison with the forward problem [51, 52]. A schematic representation of the forward and the inverse problem associated to the epidemic model is shown in Figure 3.

The inverse problem for the given SEIR-SEI model is a well-posed problem attempting to estimate a finite number of “parameters” (in the general sense of $\alpha$) in a finite dimension model. In the ways of Eq.(4), an optimization problem is implicitly established to minimize the objective function sum of squared residues between system response and experimental data.

The Trust-Region-Reflective method (TTR) is employed here to numerically approximate a solution for the inverse problem. Constraining $\alpha$ by a lower and upper bound, i.e. $\mathbf{lb} \leq \alpha \leq \mathbf{ub}$, the main idea of the method is to minimize a
Forward Problem

Mathematical Model

IC and Parameters
\( \alpha = (x_0, p) \)

Model Observable
\( \phi(\alpha, t) \)

Field Observations

\( \dot{x}(t) = f(x(t), p) \)

\( x(t_0) = x_0 \)

Inverse Problem

IC and Parameters

\( \alpha^* \)

Fitting IC and Parameters

\( y_1, y_2, \ldots, y_M \)

Field Observations

Figure 3: Schematic representation of the forward and inverse problems associated to the epidemic model.

A simpler function \( \varphi \) that reflects the behavior of \( f(\alpha) \) in a neighborhood (trust-region) around \( \alpha \). The simpler function is defined as dependent on the trial step \( s \), characterizing the Trust-Region subproblem, and its computation is optimized by restricting the subproblem to a two-dimensional subspace. The subspace is linear spanned by a multiple of the gradient \( g \) and (in the bounded case) a vector obtained in a scaled modified Newton step, used for the convergence condition \( D(\alpha)^{-2} g(\alpha) = 0 \), where \( D \) is a diagonal matrix that depends on \( \alpha, g, lb, \) and \( ub \) [53]. Finally, the trial step is found through the subproblem as

\[
 s = \arg \min_s \varphi(s) = \arg \min_s \left\{ \frac{1}{2} s^T Q s + g^T s \mid ||D s||_2^2 \leq \Delta \right\}, \quad (5)
\]

where \( \Delta \) is a scalar associated with the trust region size; \( Q \) is a matrix involving \( D \), a Jacobian matrix (also dependent on \( \alpha, g, lb, \) and \( ub \)) and an approximation of the Hessian matrix [53]. After solving Eq.(5), if \( f(\alpha + s) < f(\alpha) \) then \( \alpha \) is updated to \( \alpha + s \) and the process iterates, otherwise \( \Delta \) is decreased. In addition, a reflection step also occurs if a given step intersects a bound: the reflected step is equivalent to the original step except in the intersecting dimension, where it assumes the opposite value after reflection.

The adopted \( \alpha \) comprises all system parameters from Eq.(1) and IC, excepting \( N, R^*_h \) and \( C^i \), which were kept fixed in their nominal values; \( \phi(\alpha, t) \) is
the new cases $N_w$ system response, and the $M$ scalar observations are the data points.

The TRR algorithm also requires an initial guess for each “parameter”, identified in the next section as “TRR input”. The stopping criteria are the norm of the step and the change in the value of the objective function, with a tolerance of $10^{-7}$.

3.3. Numerical experiments for calibration

Figure 4 presents the best result for the $N_w$ system response fitting problem using the nominal parameters and IC from sections 2.3 and 3.1 as initial guesses for the TRR algorithm. The upper and lower bounds used for the parameters were set compatible with the literature suggested intervals and are presented in Table 1, along with the parameters and IC values resulted from the calibration (“TRR output”). The upper bound for $\delta$ was assumed higher than the lower bound for $\alpha_v$ to maintain consistency with the model interpretation. The minima for $S_i^h$ and $S_i^v$ were set to $0.9N$ and $0.99$, respectively, to establish a high number of susceptible individuals as is expected for the beginning of an outbreak. Also, the $lb$ and $up$ (0.999) for $S_i^v$ were motivated by noticing how variations in $S_i^h$ and $E_i^h$ below $10^{-3}$ already bring significant changes in the system response. The lack of available data for the exposed and infectious groups at the onset of the epidemic was circumvented by appointing its minimum and maximum possible values as $lb$ and $up$, i.e., $E_{v,i}$ and $I_{v,i}$ were restricted between zero and one, while $E_{h,i}$ and $I_{h,i}$ were bounded by zero and $N$. To ensure the model hypotheses of compartmentalization and constant population, two additional fitting points were defined, $\Sigma_h = S_i^h + E_i^h + I_i^h$ and $\Sigma_v = S_i^h + E_i^h + I_i^h$, which were set to match $N - R_i^h$ and 1 on Eq.(4), respectively.

![Figure 4](image1.png)  

(a) Cumulative number  
(b) New cases

Figure 4: Outbreak data (red) and calibrated model response (blue) from Table 1.

Clearly, Figure 4 is a reasonable prediction of the outbreak: the general shape of the infection evolution is attained, all parameters are within realistic possibilities, the $C(t)$ curve overshoots the data by merely 6.00%, the peak value of $N_w$ differs from the empirical data maximum by 7.87% and is only
one week off. Also, $\Sigma_h + R_h^i$ and $\Sigma_v$ amount to $N$ and 0.99 respectively, thus closely sustaining the model hypotheses. However, taking into consideration the order of magnitude of the first data point ($C_i = 8,201$) and the scale of the infection (215,319 probable cases until the 52th EW [12]), the TRR output for $I_h^i$ (253,360) is probably too high, even though there is no reference value to compare with the number of infectious individuals at the beginning of 2016, making it difficult to ascertain on a deterministic manner what is a feasible value for $I_h$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>TRR input</th>
<th>lb</th>
<th>ub</th>
<th>TRR output</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_h$</td>
<td>1/5.9</td>
<td>1/12</td>
<td>1/3</td>
<td>1/12</td>
<td>[42, 43, 44, 15, 45, 38]</td>
</tr>
<tr>
<td>$\alpha_v$</td>
<td>1/9.1</td>
<td>1/10</td>
<td>1/5</td>
<td>1/10</td>
<td>[39, 41, 38]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1/7.9</td>
<td>1/8.8</td>
<td>1/3</td>
<td>1/8.8</td>
<td>[45, 25, 38]</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1/11</td>
<td>1/21</td>
<td>1/11</td>
<td>1/16.86</td>
<td>[49, 48, 45, 47]</td>
</tr>
<tr>
<td>$\beta_h$</td>
<td>1/11.3</td>
<td>1/16.3</td>
<td>1/8</td>
<td>1/16.3</td>
<td>[38]</td>
</tr>
<tr>
<td>$\beta_v$</td>
<td>1/8.6</td>
<td>1/11.6</td>
<td>1/6.2</td>
<td>1/11.6</td>
<td>[38]</td>
</tr>
<tr>
<td>$S_h^i$</td>
<td>205,953,959</td>
<td>0.9 × N</td>
<td>N</td>
<td>205,700,000</td>
<td>——</td>
</tr>
<tr>
<td>$E_h^i$</td>
<td>8.201</td>
<td>0</td>
<td>N</td>
<td>15,089</td>
<td>——</td>
</tr>
<tr>
<td>$I_h^i$</td>
<td>8.201</td>
<td>0</td>
<td>N</td>
<td>253,360</td>
<td>——</td>
</tr>
<tr>
<td>$S_v^i$</td>
<td>0.99956</td>
<td>0.99</td>
<td>0.999</td>
<td>0.99</td>
<td>——</td>
</tr>
<tr>
<td>$E_v^i$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>——</td>
</tr>
<tr>
<td>$I_v^i$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>——</td>
</tr>
</tbody>
</table>

Table 1: TRR setup for the Figure 4 calibrated response. The values referring to parameters are in days$^{-1}$, human IC are expressed in individuals, and vector quantities in proportion.

Figure 5 allows examination of the system behavior when the $I_h^i$ value is around $C^i$, by depiction of another result to the inverse problem when the upper bound of the initial number of infectious individuals is set to 10,000. The same restriction was made over $E_h^i$ merely to simplify the analysis. Table 2 displays the resulting parameters and IC.

The model response for a 10,000 $I_h^i$ restriction also presents acceptable predictions of the general shape and numbers of the outbreak, even though it is less accurate than Figure 4 on a fitting criteria for $N_w$. For comparison, the $N_w$ peak and data maximum difference increased to 10.57% and two weeks, while the overshoot on the $C(t)$ time series actually reduced to 5.74%. Additionally, $\Sigma_v$ reached 0.999, better suiting the hypothesis.

To compare the two systems defined by the parameters and IC from each table, Figure 6 portrays their $I_h$ response. The system from Table 1 has an almost monotonically decreasing $I_h$ curve, except for a slight local maximum around the 6th EW, while the system from Table 2 presents a significant increase in the number of infectious individuals by the same time, which correspond to the weeks right before the peak infection. As stated, the lack of empirical data for the current number of infectious at each EW makes it impossible to determine what is a quantitatively reasonable prediction for $I_h$ values. But this work
Figure 5: Outbreak data (red) and calibrated model response (blue) from Table 2.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>TRR input</th>
<th>$lb$</th>
<th>$ub$</th>
<th>TRR output</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_h$</td>
<td>1/5.9</td>
<td>1/12</td>
<td>1/3</td>
<td>1/12</td>
<td>[42, 44, 44, 15, 45, 38]</td>
</tr>
<tr>
<td>$\alpha_v$</td>
<td>1/9.1</td>
<td>1/10</td>
<td>1/5</td>
<td>1/10</td>
<td>[39, 41, 38]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1/7.9</td>
<td>1/8.8</td>
<td>1/3</td>
<td>1/3</td>
<td>[45, 25, 38]</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1/11</td>
<td>1/21</td>
<td>1/11</td>
<td>1/21</td>
<td>[49, 48, 45, 47]</td>
</tr>
<tr>
<td>$\beta_h$</td>
<td>1/11.3</td>
<td>1/16.3</td>
<td>1/8</td>
<td>1/10.40</td>
<td>[38]</td>
</tr>
<tr>
<td>$\beta_v$</td>
<td>1/8.6</td>
<td>1/11.6</td>
<td>1/6.2</td>
<td>1/7.77</td>
<td>[38]</td>
</tr>
<tr>
<td>$S_h^0$</td>
<td>205,953,959</td>
<td>0.9 x N</td>
<td>N</td>
<td>205,953,534</td>
<td>————</td>
</tr>
<tr>
<td>$E_h^0$</td>
<td>8.201</td>
<td>0</td>
<td>10,000</td>
<td>6.827</td>
<td>————</td>
</tr>
<tr>
<td>$I_h^0$</td>
<td>8.201</td>
<td>0</td>
<td>10,000</td>
<td>10,000</td>
<td>————</td>
</tr>
<tr>
<td>$S_v^0$</td>
<td>0.9996</td>
<td>0.99</td>
<td>0.999</td>
<td>0.999</td>
<td>————</td>
</tr>
<tr>
<td>$E_v^0$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>0</td>
<td>1</td>
<td>$4.14 \times 10^{-4}$</td>
<td>————</td>
</tr>
<tr>
<td>$I_v^0$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>————</td>
</tr>
</tbody>
</table>

Table 2: TRR setup for the Figure 5 calibrated response. The values referring to parameters are in days$^{-1}$; human IC are expressed in individuals, and vector quantities in proportion.

assumes that a more possible scenario involves a $I_h$ time series that also follows the general shape of an epidemic curve around the weeks of maximum infection, especially considering that Figure 6(a) implicates the notion that most people were infected strictly before 2016, which does not seem to be the case suggested by the Brazilian outbreak data of probable cases of infected per EW [12] when compared to numbers available for 2015 [13]. Thus, with this qualitative criteria in mind, the system from Table 2 is selected for a further analysis over its behavior dependency to the $I_h^0$ initial condition.

Figure 7 displays the $C(t)$, $N_w$ and $I_h(t)$ responses per EW for various $I_h^0$ on the system with parameters from Table 2. To simplify the analysis, $E_h^0$ is considered equal to $I_h^0$ in each case. The remaining IC are the same from the Table. The pattern suggests that a $I_h^0$ increase on the system with this given
set of parameters continuously escalate the $C(t)$ and $N_w$ curves, eventually overshooting the data by far, and reduce the variations of $I_h(t)$ curve around its local maximum. Figure 7 allows one to make better predictions about the outbreak by analyzing the multiple possible scenarios of epidemic evolution over different values for the IC missing empirical information. Clearly, the system response in all cases is qualitative reliable in simulating the outbreak ($C(t)$ and $N_w$ shape) and can quantitatively approximate the real data values for some values of $I_h$.

4. Concluding remarks

A SEIR-SEI epidemic model to describe the dynamics of the 2016 Zika virus outbreak in Brazil is developed and calibrated in this work. Nominal quantities for the parameters are selected from the related literature concerning the Zika infection, the *Aedes aegypti* genus of mosquitoes, vector-born epidemic models and information provided by health organizations. The calibration process is done through the solution of an inverse problem with the aid of a Trust-Region-Reflective method, used to pick the best parameter values that would fit the model prediction for the number of new infectious cases per week into the disease’s empirical data. Results within realistic values for the parameters are presented, stating reasonable predictions with the curve shape similar to the outbreak evolution and proximity between the estimated peak value and data for maximum number of infected during 2016. Further analysis of the results about the lack of data for an initial condition is performed, exhibiting a range of values over which the system response keeps its quantitative reliability to a certain degree.

This work is part of a long project of modeling and prediction of epidemics related to the Zika virus in the Brazilian context [23, 24]. In upcoming studies the authors intend to take into account the uncertainties underlying the model.
parameters via Bayesian updating and employ an Active Subspace approach [54, 55, 56] to explore relevant scenarios in parametric studies.

Acknowledgments

The authors are indebted to the Brazilian agencies CNPq (National Council for Scientific and Technological Development), CAPES (Coordination for the Improvement of Higher Education Personnel) and FAPERJ (Research Support Foundation of the State of Rio de Janeiro) for the financial support given to this research. They are also grateful to João Peterson and Vinicius Lopes, both engineering students at UERJ, for the collaboration in the initial stages of this work.
References


