Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Zika virus in Brazil: calibration of an epidemic model for 2016 outbreak

Eber Dantas[1]
Michel Tosin[2]
Americo Cunha Jr[3]
NUMERICO - Núcleo de Modelagem e Experimentação Computacional
Universidade do Estado do Rio de Janeiro (UERJ)

Abstract. This work deals with the development and calibration of an epidemic model to describe the 2016 outbreak of Zika virus in Brazil. A mathematical model with 8 differential equations and 7 parameters is employed. Nominal values for the model parameters are estimated from the literature. An inverse problem associated to the model identification is formulated and solved. The calibrated model obtained presents realistic parameters and returns reasonable predictions, with the curve shape similar to the outbreak evolution and peak value close to the maximum number of new cases of infected people during 2016.

Keywords. Zika virus dynamics, nonlinear dynamics, mathematical biology, SEIR epidemic model, model calibration, system identification

1 Introduction

The Zika fever is an infectious disease caused by a homonymous flavivirus that has surged in the last two decades as multiple epidemics around the world. The widespread outbreaks of this vector-borne malady have been an international concern specially due to a suggested association with newborn microcephaly and Guillain-Barré syndrome. The Brazilian Ministry of Health confirmed 130,701 cases in Brazil by the end of 2016 [14].

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 an useful tool, which can be employed to identify effective and vulnerable aspects on disease control programs. 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 a Brazilian context [3,4], and aims to calibrate a SEIR (susceptible - exposed - infected - recovered) epidemic model with real data of the 2016 outbreak.

[1]eber.paiva@uerj.br
[2]michel.tosin@uerj.br
[3]americo@ime.uerj.br

2 Epidemic model for Zika virus dynamics

2.1 Model description

This work utilizes a variant of the Ross-Macdonald model for epidemic predictions, separating the populations into a SEIR framework: susceptible $S(t)$, those who are uncontaminated and are able to become infected; exposed $E(t)$, anyone that is carrying the pathogen but is still incapable of transmitting the disease; infectious $I(t)$, can spread the pathogen and may display symptoms; and the recovered group $R(t)$, which contains whoever is no longer infected. The following nonlinear system of ordinary differential equations governs the evolution of individuals through the SEIR groups.

$$\mathrm{d}S_h/\mathrm{d}t = -\beta_h\, S_h\, I_v\ , \tag{1}$$

$$\mathrm{d}E_h/\mathrm{d}t = \beta_h\, S_h\, I_v - \alpha_h\, E_h\ , \tag{2}$$

$$\mathrm{d}I_h/\mathrm{d}t = \alpha_h\, E_h - \gamma\, I_h\ , \tag{3}$$

$$\mathrm{d}R_h/\mathrm{d}t = \gamma\, I_h\ , \tag{4}$$

$$\mathrm{d}S_v/\mathrm{d}t = \delta - \beta_v\, S_v\, I_h/N - \delta\, S_v\ , \tag{5}$$

$$\mathrm{d}E_v/\mathrm{d}t = \beta_v\, S_v\, I_h/N - (\alpha_v + \delta)\, E_v\ , \tag{6}$$

$$\mathrm{d}I_v/\mathrm{d}t = \alpha_v\, E_v - \delta\, I_v\ , \tag{7}$$

$$\mathrm{d}C/\mathrm{d}t = \alpha_h\, E_h\ , \tag{8}$$

where the h-groups amass the number of humans at each stage of the model and the v-groups signifies proportion of vectors; N is the total human population; $1/\gamma$, the time that a human is infectious; $1/\delta$, the vector lifespan; β_h, the vector-to-human transmission rate and β_v the human-to-vector; $1/\alpha$ is the time interval an individual spends on E_h (adopted hereafter as equivalent to the time between being infected and exhibiting symptoms), h for human's and v for vector's; and $C(t)$ is the cumulative number of infectious people.

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

$$NewCases(1) = C_1\ ,\quad NewCases(w) = C_w - C_{w-1}\ ,\quad w = 2\,,3\,,4\,,\ldots,52\quad , \tag{9}$$

where C_w refers to the cumulative number of infectious for the w week.

All susceptible individuals are treated as equally capable of being infected and the recovered ones as completely immunized. Human demographical changes are not considered, and the vector population is maintained constant although variations on each vector SEIR compartment are introduced via the δ rate. The vector is regarded as a hypothetical mosquito apt to being infected or infectious throughout all its lifetime and unable to recover.

2.2 Nominal system response

The nominal values for the parameters of Eqs. (1)–(8) come from the related literature concerning the infection, the *Aedes aegypti* mosquito, and vector-borne epidemic models. Brazil had around $N = 206 \times 10^6$ people by July, 2016 [8]. The $1/\alpha_v$ is 15 days [1]; this value agrees with statistical confidence intervals (CI) presented in other works (95% CI: 4.4–17) [7]. A systematic review of the literature [10] suggests that 95% of people infected by the Zika virus who develop symptoms will do so within 11.2 days of infection (95% CI: 7.6–18.0) and will have no detectable virus in the blood by 18.9 days after infection (95% CI: 13.6–79.4). The value $1/\alpha_h = 11.2$ is compatible with the range of 3–12 days recommended in multiple sources [9,15], and thus the chosen $1/\gamma$ is $18.9 - 11.2 = 7.7$ days. As for $1/\delta$, "*the adult stage of the mosquito is considered to last an average of 11 days in the urban environment*" [12], also consistent with biological studies about the species [11] and usual life expectancy for the vector in Rio de Janeiro [6]. Finally, $1/\beta_h$ and $1/\beta_v$ have been estimated in the literature [5] as an average of 11.3 days (95% CI: 8.0–16.3) and 8.6 days (95% CI: 6.2-11.6), respectively.

Proper evaluation of the dynamic system underlying the SEIR epidemic model requires setting the initial conditions (IC). The initial time of the analysis was established as the first epidemiological week (EW) of 2016. The following are the preliminary assumptions considered in this analysis. $S_{h,i} = N - E_{h,i} - I_{h,i} - R_{h,i}$, $S_{v,i} = 1$, $E_{h,i} = I_{h,i}$, $E_{v,i} = I_{v,i}$, $C_i = I_{h,i}$, and $R_{h,i} = 0$. The value of $I_{h,i}$ is taken as 8,201, corresponding to the number of Zika fever new probable cases in Brazil on the first EW of 2016 [13]. As for $I_{v,i}$, repetitive manual estimations were tried until the resulted time series of $NewCases$ presented reasonable values compared to the real data. It became clear that the system response is very sensible to $I_{v,i}$, as slight variations in its value are required to achieve feasible results. In the process of choosing $I_{v,i}$, the matching of the $NewCases$ top value to the amplitude of infection is also a priority, since this is the main interest region for evaluation of the outbreak.

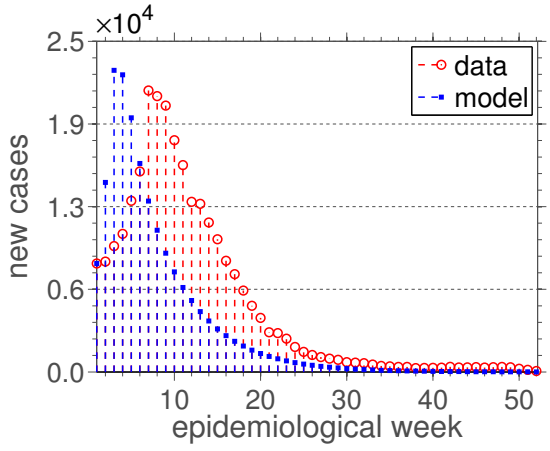

Figure 1: Model-predicted $NewCases(w)$ and number of new probable cases in each EW [13].

Viable $NewCases$ time series with the nominal parameters were possible around $I_{v,i} =$

3.5×10^{-4}. Figure 1 presents such configuration on an epidemiological week temporal domain, compared with real data of the outbreak [13] depicted by the red dots.

The given $NewCases$ model prediction clearly does not fit the infection numbers at their right time. Nevertheless, the general shape of $NewCases$ do provide qualitative information about the evolution of the infection, as well as predictions for the peak value in the same order of magnitude than that of the empirical data and its time of occurrence with a four weeks error. This qualitative agreement suggests that the model predictions may be closer to the reference values if more accurate parameter values were used.

3 Calibration of the epidemic model

3.1 Calibration method and numerical experiments

Given initial conditions and a set of parameters, represented by the pair $\boldsymbol{\alpha} = (\mathbf{x}_0, \mathbf{p})$, it is possible to compute by means of numerical integration the model response $\mathbf{x}(t)$ of the continuous-time dynamical system of section 2, from which a scalar observable $\phi(\boldsymbol{\alpha}, t)$ is obtained. In this manner, the calibration of the model consists in finding a set of parameters and initial conditions $\boldsymbol{\alpha}^*$ such that

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha}) = \arg\min_{\boldsymbol{\alpha}} \left\{ \sum_{n=1}^{M} \Big| y_n - \phi\left(\boldsymbol{\alpha}, t_n\right) \Big|^2 \right\}, \tag{10}$$

where $y_1, y_2, \ldots, y_M$ are M system observations (reference data) assigned to the $\{t_n\}_{n=1}^{M}$ time instants. This is the associated *inverse problem*.

The Trust-Region-Reflective method (TTR) is employed here to numerically approximate a solution for the inverse problem. Constraining $\boldsymbol{\alpha}$ by a lower and upper bound, i.e. $\mathbf{lb} \leq \boldsymbol{\alpha} \leq \mathbf{ub}$, the main idea of the method is to minimize a simpler function φ that reflects the behavior of $f(\boldsymbol{\alpha})$ in a neighborhood (trust-region) around $\boldsymbol{\alpha}$. The simpler function is defined as dependent on the trial step $\mathbf{s}$, 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 $\mathbf{g}$ and (in the bounded case) a vector obtained in a scaled modified Newton step, used for the convergence condition $D(\boldsymbol{\alpha})^{-2}\,\mathbf{g}(\boldsymbol{\alpha}) = 0$, where D is a diagonal matrix that depends on $\boldsymbol{\alpha}$, $\mathbf{g}$, $\mathbf{lb}$, and $\mathbf{ub}$ [2]. Finally, the trial step is found through the subproblem as

$$\mathbf{s} = \arg\min_{\mathbf{s}} \varphi(\mathbf{s}) = \arg\min_{\mathbf{s}} \left\{ \frac{1}{2}\mathbf{s}^{\mathrm{T}} Q \mathbf{s} + \mathbf{g}^{\mathrm{T}}\mathbf{s} \;\mid\; ||D\,\mathbf{s}||_2 \leq \Delta \right\}, \tag{11}$$

where Q is a matrix involving a approximation of the Hessian matrix, D and a jacobian matrix that also depends on $\boldsymbol{\alpha}$, $\mathbf{g}$, $\mathbf{lb}$, and $\mathbf{ub}$; Δ is a scalar associated with the trust region size [2].

After optimize (11), if $f(\boldsymbol{\alpha} + \mathbf{s}) < f(\boldsymbol{\alpha})$ then $\boldsymbol{\alpha}$ is updated to $\boldsymbol{\alpha} + \mathbf{s}$ and the process iterates, otherwise Δ is decreased. In addition, a reflection step also occurs if a given step intersects a bound: the reflected step is equal to the original step except in the intersecting dimension, where it assumes the opposite value after reflection.

The adopted $\boldsymbol{\alpha}$ for the SEIR model calibration in this work includes all system parameters and IC, excepting N, $R_{h,i}$ and C_i, which were kept fixed in their values of section 2.2.

3.2 Calibration results

Figure 2 presents the best result for the $NewCases$ system response fitting problem using the nominal parameters and IC from section 2.2 as initial guesses for the TRR. The upper and lower bounds used for the parameters were set compatible with the literature suggested intervals. The minimum for $S_{h,i}$ and $S_{v,i}$ were $0.9\,N$ and 0.9, to establish a high number of susceptible individuals, as is expected for the beginning of a outbreak. $E_{v,i}$ and $I_{v,i}$ were restricted between zero and one, while $E_{h,i}$ and $I_{h,i}$ were bounded to a maximum of $100\,C_i$. The resulting values that generate the time series are displayed in Table 1.

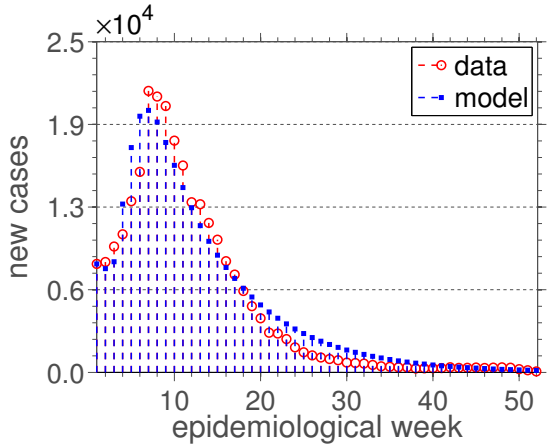

Figure 2: Calibrated $NewCases(w)$ from Table 1 and new probable cases in each EW.

It is clear in Figure 2 that the $NewCases$ system response is a reasonable prediction of the outbreak: the general shape of the infection evolution is attained, the peak value differs from the empirical data maximum only by 6.9% and are in the same week, and all parameters are within realistic possibilities. However, improvements can be made on the simulation perspective of the analysis, since $I_{h,i} = 324{,}753$ individuals is a high estimation.

Another result to the inverse problem is presented in Figure 3, considering again the fitting of the $NewCases$. This time, after defining $Sum_h = S_{h,i} + E_{h,i} + I_{h,i}$ and $Sum_v = S_{v,i} + E_{v,i} + I_{v,i}$, another two data points are provided for the TRR algorithm: $Sum_h = N - R_{h,i}$ and $Sum_v = 1$. Thus, this setup of the method tries to ensure the compartmentalization hypothesis stays true through the TRR choosing of $\boldsymbol{\alpha}$. The upper bound of $I_{h,i}$ was diminished to $20,000$ individuals and the $S_{v,i}$ lower limit set to 0.99. The resulting values that generate this second calibrated time series are in Table 2.

The $NewCases$ system response in Figure 3 is still a reasonable prediction of the general shape and numbers of the outbreak, even though it is less accurate than Figure 2 on a fitting criteria (for comparison, the peak and data maximum differ by 10.6%). Nevertheless, significance of this result resides on a completely reasonable set of parameters and initial conditions: 99.98% of the human population and 99.94% of the vectors are withing the compartmentalization hypothesis of the model, and the value of $I_{h,i}$ is close to C_i.

Table 1: TRR results for Figure 2. Parameters: days. Human IC: individuals.

$\beta_h^{-1} = 16.3$	$\alpha_h^{-1} = 12.0$	$\gamma^{-1} = 12.0$	$\beta_v^{-1} = 11.6$	$\alpha_v^{-1} = 15.0$	$\delta^{-1} = 15.27$
$S_{h,i}$	$E_{h,i}$	$I_{h,i}$	$S_{v,i}$	$E_{v,i}$	$I_{v,i}$
185.4×10^6	16,664	324,753	0.9	2.22×10^{-14}	2.22×10^{-14}

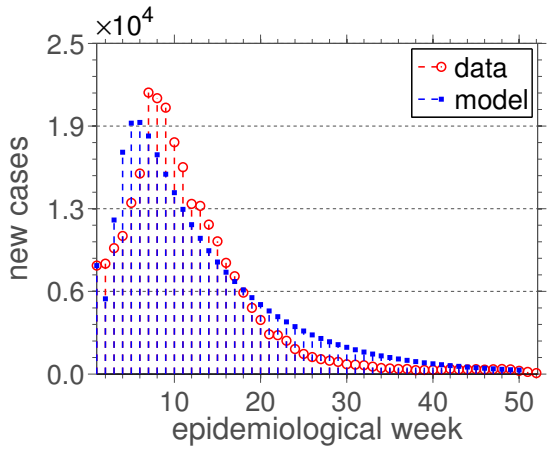

Figure 3: Calibrated $NewCases(w)$ from Table 2 and new probable cases in each EW.

Table 2: TRR results for Figure 2. Parameters: days. Human IC: individuals.

$\beta_h^{-1} = 9.9$	$\alpha_h^{-1} = 12.0$	$\gamma^{-1} = 4.0$	$\beta_v^{-1} = 11.6$	$\alpha_v^{-1} = 12.0$	$\delta^{-1} = 22.0$
$S_{h,i}$	$E_{h,i}$	$I_{h,i}$	$S_{v,i}$	$E_{v,i}$	$I_{v,i}$
205.9×10^6	7,859	20,000	0.998	4.05×10^{-4}	2.49×10^{-14}

4 Final remarks

A SEIR epidemic model to describe the dynamics of the 2016 Zika virus outbreak in Brazil is developed and calibrated in this work. Nominal parameter quantities are selected from the related literature. The calibration is done through the solution of an inverse problem with a Trust-Region-Reflective method, used to pick the best parameter and IC 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. A second result also including suitable IC is achieved with a less accurate fitting.

Acknowledgments

The authors are indebted to the Brazilian agencies CNPq, CAPES, and FAPERJ for the financial support given to this research. They are also grateful to João Peterson and Vinícius Lopes, both engineering students at UERJ, for the collaboration on early stages of this work.

References

[1] J. P. T. Boorman and J. S. Porterfield. A simple technique for infection of mosquitoes with viruses transmission of Zika Virus, *Trans. R. Soc. Trop. Med. Hyg.*, OUP, 50:238–242, 1956. DOI: 10.1016/0035-9203(56)90029-3.

[2] T. F. Coleman, Y. Li., An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds, *SIOPT*, SIAM, 6:418–445, 1996.

[3] E. Dantas, M. Tosin, J. Peterson, V. Lopes, A. Cunha Jr. A mathematical analysis about Zika virus outbreak in Rio de Janeiro. In *Proceedings of the* 4*th Conference of Computational Interdisciplinary Sciences (CCIS 2016)*, SP, Brasil, 2016.

[4] E. Dantas, M. Tosin, A. Cunha Jr. Calibration of a SEIR epidemic model to describe Zika virus outbreak in Brazil, 2017. https://hal.archives-ouvertes.fr/hal-01456776v2

[5] N. M. Ferguson, et al. Countering the Zika epidemic in Latin America, *Science*, AAAS, 353:353–354, 2016. DOI: 10.1126/science.aag0219.

[6] R. M. de Freitas, C. T. Codeço and R. L. de Oliveira. Daily survival rates and dispersal of Aedes aegypti females in Rio de Janeiro, Brazil, *Am. J. Trop. Med. Hyg.*, ASTMH, 76:659–665, 2007. PMID: 17426166.

[7] S. Funk, et al. Comparative analysis of Dengue and Zika outbreaks reveals differences by setting and virus, *PLoS Negl. Trop. Dis.*, PLoS, 10(12):e0005173, 2016.

[8] IBGE.*D.O.U*, Imprensa Nacional, number 167, section 1, pages 47–65, 2016.

[9] S. Ioos, et al. Current Zika virus epidemiology and recent epidemics, *Med. Mal. Infect.*, Elsevier, 44:302–307, 2014. DOI: 0.1016/j.medmal.2014.04.008.

[10] J. T. Lessler, et al. Times to key events in Zika virus infection and implications for blood donation: a systematic review, *Bull. World. Health.*, WHO, 94:841–849, 2016. DOI: 10.2471/BLT.16.174540.

[11] M. J. Nelson. *Aedes Aegypti: Biology and Ecology*. PAHO, Washington, D.C., 1986.

[12] M. Otero, H. G. Solari and N. Schweigmann. A stochastic population dynamics model for Aedes Aegypti: formulation and application to a city with temperate climate, *Bull. Math. Biol.*, Springer, 68:1945–1974, 2006. DOI: 10.1007/s11538-006-9067-y.

[13] Ministério da Saúde. Pedido 25820000408201713, *e-SIC* <`https://goo.gl/7vAx1a`> (Jan 26,2017)

[14] Secretaria de Vigilância em Saúde. Zika virus - Boletim Epidemiológico, v. 48, n. 3, 2017. ISSN: 2358-9450. <`https://goo.gl/QO3hsy`> (Mar 10, 2017)

[15] G. Valentine, L. Marquez and M. Pammi. Zika virus-associated microcephaly and eye lesions in the newborn, *J. Pediatric. Infect. Dis. Soc.*, OUP, 5:323–328, 2016. DOI: 10.1093/jpids/piw037.