Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, v. 6, n. 1, 2018.

Trabalho apresentado no XXXVII CNMAC, S.J. dos Campos - SP, 2017.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Zika virus in Brazil: calibration of a epidemic model for the 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 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 confirms that the first autochthonous transmission of Zika virus in Brazil happened around April, 2015, and has registered 130,701 confirmed cases 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 [2,3], and aims at calibrating a SEIR 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.

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 (main vector for Zika in Brazil), vector-borne epidemic models and reports by health and government agencies. Brazil had around $N = 206 \times 10^6$ people by July, 2016 [7]. 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) [6]. A systematic review of the literature [9] 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 [8, 15]. Considering the assumption that the infectiousness in Zika infection ends 1.5–2 days before the virus becomes undetectable [4,6], the chosen $1/\gamma$ is $18.9 - 1.5 = 17.4$. As for $1/\delta$, "*the adult stage of the mosquito is considered to last an average of eleven 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 [5]. Finally, $1/\beta_h$ and $1/\beta_v$ have been estimated in the literature [4] 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 values of its variables at $t = 0$, established as the first epidemiological week (EW) of 2016. The following are the assumptions considered in this analysis. $S_h(0) = N$, $S_v(0) = 1$, $E_h(0) = I_h(0)$, $E_v(0) = I_v(0)$, $C(0) = I_h(0)$, and $R_h(0) = 0$. The value of $I_h(0)$ is taken as 4,272, corresponding to the number of Zika fever confirmed cases in Brazil on the first EW of 2016 [13]. As for $I_v(0)$, repetitive manual estimations were tried until the resulted time series of I_h presented reasonable values compared to the real data. It became clear that the system response is very sensible to $I_v(0)$, as slight variations in its value are required to achieve feasible results. In the process of choosing its value, the matching of the I_h curve's peak to the amplitude of infection is also a priority, since this is the main interest region for evaluation of the outbreak. Viable I_h curves with the nominal parameters were possible around $I_v(0) = 6.5 \times 10^{-5}$. Figure 1 presents such configuration on a epidemiological week temporal domain consisting of one to fifty-two weeks (7 to 365 days), compared with real data of the outbreak [13] depicted by the red dots.

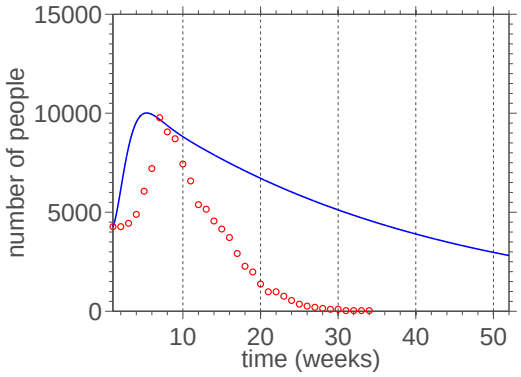

Figure 1: Model-predicted $I_h(t)$ (curve) and number of confirmed cases in each EW [13] (dots).

The given I_h curve clearly overestimates the infection numbers. Nevertheless, the general shape of I_h 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 less than two weeks error. This qualitative agreement suggests that the model predictions may be closer to the reference values if more accurate parameter values were used.

DOI: 10.5540/03.2018.006.01.0460

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 $(\mathbf{x}_0, \mathbf{p})$, 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 $\phi(\mathbf{x}_0, \mathbf{p}, t)$ is obtained. In this manner, the calibration of the model consists in finding a set of parameters $\mathbf{p}^*$ such that

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} \left\{ \sum_{n=1}^{M} \Big| y_n - \phi\left(\mathbf{x}_0, \mathbf{p}, t_n\right) \Big|^2 \right\}, \tag{9}$$

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

The method of Levenberg-Marquardt (LM) is employed here to numerically approximate a solution for the inverse problem. Its basis resides in the linear approximation $\phi(\mathbf{x}_0, \mathbf{p} + \mathbf{h}, t) \approx \phi(\mathbf{x}_0, \mathbf{p}, t) + \mathbf{J}\,\mathbf{h}$, where $\mathbf{J} = \partial\phi/\partial\mathbf{p}$ is the Jacobian matrix at $\mathbf{p}$, and $\mathbf{h}$ a small perturbation. Starting at an initial guess, the method produces a series of vectors vectors $\mathbf{p}^{(1)}$, $\mathbf{p}^{(2)}$, ..., that converge towards a local minimizer $\mathbf{p}^*$ for ϕ. The rth iteration is defined by $(\mathbf{A}^{(r)} + \lambda^{(r)}\mathbf{I})\,\mathbf{h}^{(r)} = \mathbf{g}^{(r)}$ [10], where $\mathbf{A} = \mathbf{J}^T\mathbf{J}$, $\mathbf{g} = \mathbf{J}^T\left(\mathbf{y} - \phi(\mathbf{x}_0, \mathbf{p}, t) - \mathbf{J}\,\mathbf{h}\right)$, and the scalar λ is a Lagrange multiplier that controls both the magnitude and direction of the step size correction $\mathbf{h}$. A λ equal to zero implies that the direction of $\mathbf{h}$ is identical to that of the Gauss-Newton method, and as $\lambda \to \infty$, $\mathbf{h}$ tends towards the steepest descent method direction with its magnitude tending to zero. Thus, λ can be controlled to ensure descent even when second-order terms are encountered, which would otherwise restrict the efficiency of the Gauss-Newton method. The parameter vector at each step, $\mathbf{p}^{(r+1)} = \mathbf{p}^{(r)} + \mathbf{h}^{(r)}$, leads to a new sum of squares in Eq.(9), and it is essential to select λ such that this sum is strictly smaller than the previous one. A sufficiently large λ that ensures this condition is met always exists, unless $\mathbf{p}^{(r)}$ is already at a minimum [10].

Variations of a single parameter via the LM algorithm, while maintaining the others constant, revealed that the $I_h(t)$ response of the model was significantly sensible to the β_h and β_v rates, being largely more affected by β_h. Next, a two-varying-parameters attempt was conducted: most combinations did not bring satisfactory results, to the extent that some pairs of parameters could not even be computed in the time dedicated to the analysis, e.g (α_h, α_v), probably because of inefficient initial guesses or high computational cost. Besides, β_h and β_v seemed to control the quality changing of the considered curve, meaning the accompanying parameter would vary relatively less, proving the pair to be the best one for fitting purposes. Three-way-varying attempts were made, but usually would result not computable or giving in to the referred control parameters.

3.2 Calibration results

Figure 2 presents the best result for the I_h curve fitting problem using the nominal parameters, obtained by singly varying β_v via the LM algorithm. The β_v value for the

initial guess, the parameters held constant and the initial conditions are the same as described in section 2.2. The resulting $1/\beta_v$ that graphs the curve is 15.5 days.

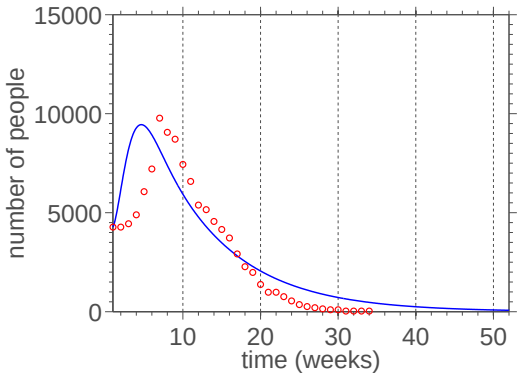

Figure 2: Calibrated time series of $I_h(t)$ using the nominal parameters.

It is clear in Figure 2 that the system response I_h is a reasonable prediction of the outbreak: the general shape of the infection evolution is attained, the curve's peak and empirical data maximum value differ only by a couple hundred of individuals, and all parameters and initial conditions are within realistic possibilities.

Another result to the inverse problem is presented, considering again the fitting of the I_h curve. The chosen values for this second set of parameters are the product of comparing the empirical data with multiple iterations of the numerical strategy, but lacking the caution for the possibility of unrealistic measures for the parameters. The initial conditions for the numerical integration in this analysis follow the same assumptions presented in section 2.2, excepting the $I_v(0) = E_v(0)$ hypothesis, since the values of these are now the result of an additional heuristic process of calibration: successive applications of the LM algorithm were performed while manually changing the values of $I_v(0)$ and $E_v(0)$ at each application, searching for the best fit of the curve's peak to the field data. Figure 3 presents the most satisfactory result obtained through this analysis, utilizing a β_h variation in the LM method. The parameters maintained constant and the β_h initial guess used in this calibration process to graph Figure 3 are summarized in Table 1, along with the values of $I_v(0)$ and $E_v(0)$. The resulting β_h after the LM algorithm is $0.0127\,\text{days}^{-1}$.

Table 1: Parameters, $I_v(0)$ and $E_v(0)$ used for Figure 3. The β_h value is the initial guess for the LM algorithm. The remaining values are the same ones from section 2.2.

	α_h	α_v	γ	δ	β_h	β_v	$I_v(0)$	$E_v(0)$
value	1/12	1/17	1/8.8	1/25	1/16.3	1/11.6	15×10^{-5}	135×10^{-5}
unit	days^{-1}	days^{-1}	days^{-1}	days^{-1}	days^{-1}	days^{-1}	—	—

The curve in Figure 3 presents a better calibration of the model according to the empirical data, since the peak time of the I_h curve is significant closer to the epidemiological week that registered the maximum number of infected people. However, this result comes at the cost of physical meaning in the parameters, because $1/\beta_h = 78.7$ days for the time

between a mosquito being infected and it infecting a human is certainly unrealistic.

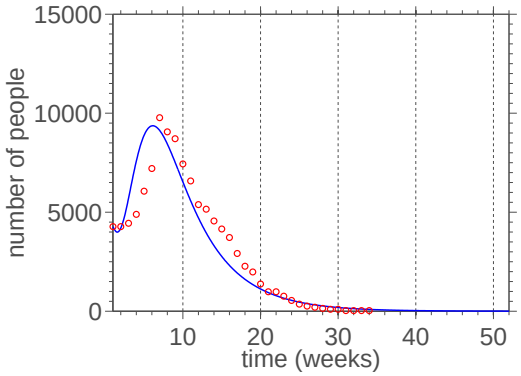

Figure 3: Calibrated time series of infectious number of humans using parameters from Table 1.

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 process is done through the solution of an inverse problem with the aid of the Levenberg-Marquardt method, used to pick the best parameter values that would fit the curve "number of infectious people per week" into the disease's empirical data, thus calibrating the model. 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. Improved fitting is also achieved via convenient choice of parameters and initial conditions during the numerical and heuristic process regarded in the analysis, but at the expense of physical meaning of such parameters.

This work is only the first step in a long project of modeling and prediction of epidemics related to the Zika virus in the Brazilian context. In future works, the authors intend to analyze the efficacy of other epidemic models (e.g. SIR, MSIR, etc), and improve the calibration process by means of a Bayesian updating rule to attack the inverse problem, turning it more robust by taking into account the uncertainties underlying the model and its parameters. The use of statistical methods to choose the most appropriate model within a set of validated ones is also part of the plans.

Acknowledgments

CNPq, CAPES, and FAPERJ, for the financial support. UERJ engineering students João Peterson and Vinícius Lopes, for 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] 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 4th Conference of Computational Interdisciplinary Sciences (CCIS 2016)*, São José dos Campos, SP, Brasil, 2016.

[3] E. Dantas, M. Tosin, A. Cunha Jr., Calibration of a SEIR epidemic model to describe Zika virus outbreak in Brazil, 2017. <https://goo.gl/NamHPU> (Mar 10, 2017)

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

[5] 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.

[6] 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. DOI: 10.1371/journal.pntd.0005173.

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

[8] 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.

[9] 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.

[10] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, SIAM, 11:431–441, 1963. DOI: 10.1137/0111030.

[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] PAHO, WHO. Zika - Epidemiological Report Brazil, March, 2017. <`https://goo.gl/StOsHC`> (Mar 10, 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.

