# An uncertainty quantification framework for a Zika virus epidemic model

Eber Dantas, Michel Tosin, Americo Cunha Jr

# An uncertainty quantification framework for a Zika virus epidemic model

Eber Dantas[1], Michel Tosin[2] and Americo Cunha Jr.[3]

Universidade do Estado do Rio de Janeiro – UERJ, Rio de Janeiro – RJ, Brazil

Received on February 02, 2019 / accepted on *****, 2019

**Abstract**

Uncertainty quantification is an important procedure when dealing with errors and discrepancies that are present in any modeling effort. This work presents a consistent uncertainty quantification framework for an epidemiological dynamical system, which is able to construct robust descriptions given a calibrated model. Since arbitrary choices of distributions for the input parameters can provide biased estimates and results, the maximum entropy principle is employed in the construction of the stochastic model to infer the most possibly unbiased probability density functions affected by the lack of information. The framework is applied on a SEIR-SEI compartmental system for the Brazilian Zika virus outbreak to study a stochastic scenario.

**Keywords**: Zika Virus, nonlinear dynamics, uncertainty quantification, maximum entropy principle, epidemic model.

## 1. Introduction

A primordial step in the application of mathematical models for the study of natural and artificial phenomena is the validation of such models through several calibration process with independent data sets. Unfortunately, that process can be poorly influenced by data that underestimates the reality or due to inaccurate values of the parameters: small variations on the inputs, in some cases, can bring about huge changes in the model response. Propagation of uncertainties introduced in the parameters allows one to observe if the model provides robust results when confronted with multiple sources of errors and discrepancies [1, 2].

This work presents an uncertainty quantification (UQ) framework, and showcase its use for an epidemiological compartmental model previously

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

calibrated to describe the outbreak of the Zika virus in Brazil [3]. In order to do a refined study while considering the more influential parameters for the response, uncertainties are propagated only for the more sensible inputs, which were determined and analyzed in another previous work [4]. The lack of knowledge regarding the input distribution introduces bias when naive choices are made. The maximum entropy principle [2] is detailed on the application of selecting the parameter's distribution consistently, picking the less biased one available.

## 2. Computational model

The SEIR-SEI nonlinear dynamical system outlined in Figure 1 was previously calibrated for the Brazilian context of Zika infection with values from the literature [3]. This compartmental model divides the human and vector populations into four groups according to their current state (time $t$) in relation to the Zika fever: susceptible for infection, $S(t)$; exposed to the virus but unable to transmit it, $E(t)$; infectious, $I(t)$, which are infected and capable of transmitting the disease; and recovered humans, $R_h(t)$. Demographic changes are only considered for the vector, with both populations constant.

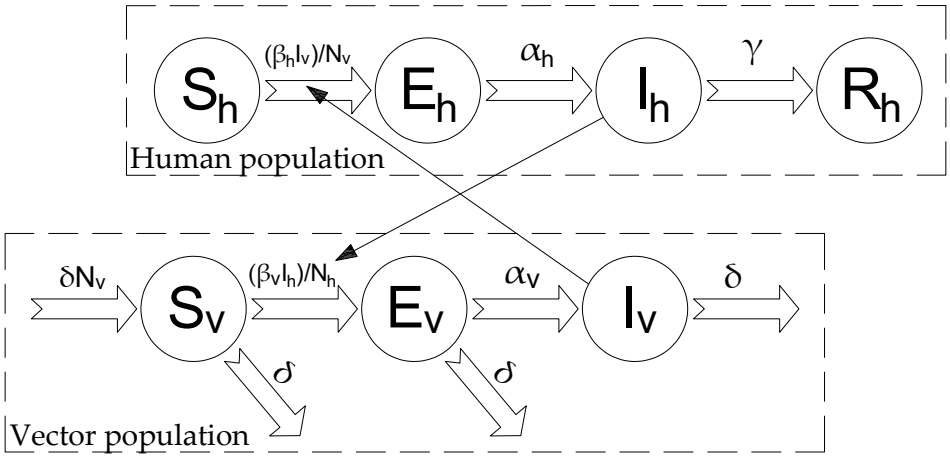


Figure 1: Diagram of the SEIR-SEI system dynamics.

The transmission mechanism embedded in the continuous-time dynamical system is the cross-infection, as expressed in the system of differential equations

$$
\begin{aligned}
\frac{\mathrm{d}S_h}{\mathrm{d}t} &= -\beta_h\, S_h\, \frac{I_v}{N_v}\,, & \frac{\mathrm{d}S_v}{\mathrm{d}t} &= \delta\, N_v - \beta_v\, S_v\, \frac{I_h}{N_h} - \delta\, S_v\,, \\
\frac{\mathrm{d}E_h}{\mathrm{d}t} &= \beta_h\, S_h\, \frac{I_v}{N_v} - \alpha_h\, E_h\,, & \frac{\mathrm{d}E_v}{\mathrm{d}t} &= \beta_v\, S_v\, \frac{I_h}{N_h} - (\alpha_v + \delta)\, E_v\,, \\
\frac{\mathrm{d}I_h}{\mathrm{d}t} &= \alpha_h\, E_h - \gamma\, I_h\,, & \frac{\mathrm{d}I_v}{\mathrm{d}t} &= \alpha_v\, E_v - \delta\, I_v\,, \\
\frac{\mathrm{d}R_h}{\mathrm{d}t} &= \gamma\, I_h\,, & \frac{\mathrm{d}C}{\mathrm{d}t} &= \alpha_h\, E_h\,,
\end{aligned}
\tag{1}
$$

where $1/\alpha$ represents the incubation period, $1/\delta$ is the vector lifespan, $1/\gamma$ the human infection period, $\beta_h$ and $\beta_v$ the transmission rates. The total human population is $N_h = 206 \times 10^6$, the Brazilian population in 2016, and $N_v = 1$ since the vector population is treated in proportions. The cumulative number of infectious is collected for each time on $C(t)$, in the last equation.

The vector of stated obtained from Eqs.(1) reports how humans and vectors are segmented for each time t. The quantities of interest (QoI) are the cumulative number of infectious and the new cases per epidemiological week (EW), computed from

$$
\begin{aligned}
C(t) &= \int_{t_0}^{t} \alpha_h\, E_h\, \mathrm{d}t\,, \\
\mathcal{N} &= C_w - C_{w-1}\,, \quad \mathcal{N}_1 = C_1, \quad w = 2, \cdots, 52\,,
\end{aligned}
\tag{2}
$$

where $C_w$ is the value of $C$ in the $w$-th EW. The QoI in this way are quantities associated to the epidemiological concepts of prevalence and incidence, which can be compared to real data about the Brazilian Zika infection of 2016 assembled by the Ministry of Health [5].

The computational model uses the input parameters ($\beta_h$, $\alpha_h$, $\gamma$, $\beta_v$, $\alpha_v$, $\delta$) in Eqs.(1) and then obtain the QoIs through Eqs.(2) to output $C(t)$ and $N_w$.

## 3. Uncertainty Quantification

The dynamical system from the previous section provides QoI that can be seen as a random quantity due to the variability (randomness) in the system parameters. In this context, the mathematical model defined by Eqs.(1) and Eqs.(2) can be thought abstractly as an operator $\mathcal{M}$, which maps an input (random) vector $\mathbf{X}$ that lumps the system random parameters, into an output (random) scalar $Y$, representing a generic QoI, i.e.,

$$Y = \mathcal{M}(\mathbf{X}). \tag{3}$$

To deal with the uncertainties of the problem, a consistent probabilistic model must be constructed. In general, for lack of information, the joint-distribution of $\mathbf{X}$ can not be reliably specified via non-parametric statistics. In this way, a conservative procedure aims to estimate the least biased distribution, while using the few information known about it. The maximum entropy principle (MaxEnt) is formalized over this premise. The joint-PDF of $\mathbf{X}$, denoted by $p_{\mathbf{X}}(\boldsymbol{x})$, is obtained by maximizing the Shannon Entropy

$$\mathcal{E}(p_{\mathbf{X}}) = -\int_{\mathbb{R}} p_{\mathbf{X}}(\boldsymbol{x})\, \ln p_{\mathbf{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}\,, \tag{4}$$

with the restrictions

$$\int_{\mathbb{R}} \boldsymbol{g}(\boldsymbol{x})\, p_{\mathbf{X}}(\boldsymbol{x})\, \mathrm{d}\mathbf{x} = \mathbf{b}\ , \tag{5}$$

where $\boldsymbol{g}(\boldsymbol{x}) = \big(1, g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_u(\boldsymbol{x})\big)$ is a $\mathbb{R}^{u+1}$-valued function defined on the support $I_{\mathbf{X}}$, that enforces certain statistical properties $\mathbf{b} \in \mathbb{R}^{u+1}$ along with the PDF normalization condition. The general solution for this optimization problem can be written with Lagrange multipliers [2] as

$$p_{\mathbf{X}}(\boldsymbol{x}) = \mathbb{1}_{I_{\mathbf{X}}}\, \exp\big(- < \boldsymbol{\lambda}, \boldsymbol{g}(\boldsymbol{x}) >\big)\,, \tag{6}$$

with $\boldsymbol{\lambda}$ being the $u$-dimensional vector of Lagrange multipliers and $< \cdot, \cdot >$ denoting the dot product. The multipliers are found by solving the nonlinear system of equations that appear when substituting Eq.(6) in Eq.(5) to guarantee $\mathbf{b}$ [2].

In terms of stochastic calculations, once the input random vector $\mathbf{X}$ distribution is known, the distribution of output $Y$ is determined a posteriori, via Monte Carlo simulation [1, 2].

## 4. Results

Global sensitivity measures developed in a parallel work of the authors [4] suggest that $\beta_h$ and $\delta$ are the most important inputs when studying how the system response around EW 7 (peak of the outbreak). A stochastic scenario is investigated in Figure 2, where 95%-confidence bands are presented for the system QoI when introducing dispersion in the MaxEnt probability density function. For the construction of this PDF, the parameter's support bounds came from [3], and the mean value was considered the set that provided the calibrated response in [4]. A coefficient of variation ($CV$) of 10% for $\beta_h$ and 15% for $\delta$ is assumed to observe its effect on the output.

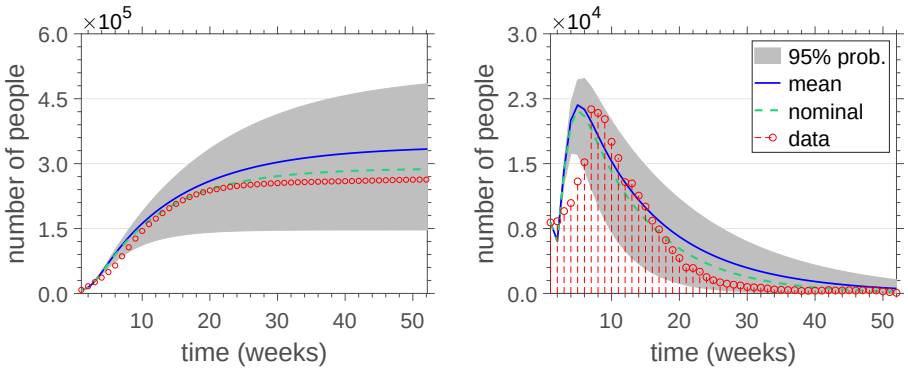


Figure 2: The 95%-confidence band for the cumulative number $C(t)$ (left) and new cases $\mathcal{N}_w$ (right).

The graphs in Figure 2 allow to contemplate the robustness of the model in comparison to the epidemic data, showing how the system response keeps its general shape and follows the outbreak evolution around the peak infection until the end, even under some variability in the parameters. Another information regards the initial time, where it is apparent that these parameters have little control in the QoI at early weeks, due to small width of the confidence bands. This procedure exemplifies how the UQ-framework can be employed to study different kinds of uncertainties subjected over the model and made in a mathematical consistent manner.

### 5. Concluding remarks

An uncertainty quantification framework that uses the the Maximum Entropy Principle to estimate the parameter's distribution was used, evaluating the robustness of a SEIR-SEI calibrated model for the context of the 2016 Zika infection in Brazil. A stochastic scenario were tested, estimating the most unbiased input distributions, promoting propagation of uncertainties through the system and calculating confidence interval for the response. In future works, the authors intend to apply the uncertainty quantification framework in more complex models and also increment the analysis by employing Bayesian updating.

**Acknowledgments**. The authors acknowledge the support given to this research by the funding agencies Carlos Chagas Filho Research Foundation of Rio de Janeiro State (FAPERJ) under grants E-26/010.002.178/2015 and E-26/010.000.805/2018, and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

## References

[1] Smith, R.C. Uncertainty Quantification: Theory, Implementation, and Applications, 2nd Edition, 2014, SIAM.

[2] Soize, C. Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering, 1st Edition, 2017, Springer.

[3] Dantas, E., Tosin, M., Cunha Jr, A. Calibration of a SEIR–SEI epidemic model to describe the Zika virus outbreak in Brazil, Applied Mathematics and Computation 338: 249-259, 2018. doi:10.1016/j.amc.2018.06.024

[4] Dantas, E., Tosin, M., Cunha Jr, A. Uncertainty quantification in the nonlinear dynamics of Zika virus, Under review, 2019. https://hal.archives-ouvertes.fr/hal-02005320v1

[5] Ministério da Saúde. Obtenção de número de casos confirmados de zika, por município e semana epidemiológica, `http://www.consultaesic.cgu.gov.br/busca/dados/Lists/Pedido/Item/displayifs.aspx?List=0c839f31-47d7-4485-ab65-ab0cee9cf8fe&ID=537263&Web=88cc5f44-8cfe-4964-8ff4-376b5ebb3bef`, 2017. Accessed: 01/19/2019