



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Original software publication

ARBO: Arbovirus modeling and uncertainty quantification toolbox Michel Tosin <sup>a</sup>, Eber Dantas <sup>b</sup>, Americo Cunha Jr <sup>a,\*</sup>, Rebecca E. Morrison <sup>c</sup><sup>a</sup> Rio de Janeiro State University, Rio de Janeiro, Brazil<sup>b</sup> Federal University of Rio de Janeiro, Rio de Janeiro, Brazil<sup>c</sup> University of Colorado Boulder, Boulder, USA

## ARTICLE INFO

## Keywords:

Arbovirus  
 Mathematical epidemiology  
 Model calibration  
 Uncertainty quantification  
 Model discrepancy

## ABSTRACT

The ongoing pandemic of COVID-19 has highlighted the importance of mathematical tools to understand and predict outbreaks of severe infectious diseases, including arboviruses such as Zika. To this end, we introduce ARBO, a package for simulation and analysis of arbovirus nonlinear dynamics. The implementation follows a minimalist style, and is intuitive and extensible to many settings of vector-borne disease outbreaks. This paper outlines the main tools that compose ARBO, discusses how recent research works about the Brazilian Zika outbreak have explored the package's capabilities, and describes its potential impact for future works on mathematical epidemiology.

## Code metadata

Current code version	v1.0
Permanent link to code/repository used for this code version	<a href="https://github.com/SoftwareImpacts/SIMPAC-2021-201">https://github.com/SoftwareImpacts/SIMPAC-2021-201</a>
Permanent link to Reproducible Capsule	<a href="https://codeocean.com/capsule/8169007/tree/v4">https://codeocean.com/capsule/8169007/tree/v4</a>
Legal Code License	MIT license
Code versioning system used	git
Software code languages, tools, and services used	Matlab and C++
Compilation requirements, operating environments & dependencies	UQLab for Matlab; QUESO and Eigen3 for C++
If available Link to developer documentation/manual	<a href="https://github.com/amicocunhajr/ARBO">https://github.com/amicocunhajr/ARBO</a>
Support email for questions	<a href="mailto:michel.a.t.c.137@gmail.com">michel.a.t.c.137@gmail.com</a>

## 1. Introduction

Arboviruses have been emerging in different parts of the world over the past several decades. Tropical countries seem to be more susceptible to these diseases due to weather favoring the proliferation of mosquitoes [1]. The last update by the Pan American Health Organization reported that more than 1.6M cases of arboviral diseases occurred in the Americas between the first and 42nd epidemiological weeks of 2020 [2]. Even though COVID-19 has of late dominated the agendas of global health institutions, the grave and harmful effects of arboviruses should not be neglected [3,4]. Indeed, after the connection between Zika infection and microcephaly in newborns was revealed, the World Health Organization declared the disease a global medical emergency

in 2016. As new outbreaks of arboviruses occur, new medical concerns will surely arise.

To simulate infectious diseases transmitted by mosquitoes, double population compartmental models represent the cross-infection dynamics between humans and vectors [5–8]. Widely explored in the literature, this kind of homogeneous model provides a good balance between interpretability and simplicity. In this vein, the ARBO package was developed to model and analyze the Zika virus outbreak that occurred in Brazil in 2016 [9], but can be easily adapted to guide studies in other regions and for other arboviruses such as Dengue or Chikungunya. Moreover, the package is written and organized in a didactic form, allowing it to also be used for educational purposes.

The code (and data) in this article has been certified as Reproducible by Code Ocean: (<https://codeocean.com/>). More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals>.

\* Corresponding author.

E-mail addresses: [michel.tosin@uerj.br](mailto:michel.tosin@uerj.br) (M. Tosin), [eberdantas@mecanica.coppe.ufrj.br](mailto:eberdantas@mecanica.coppe.ufrj.br) (E. Dantas), [amico.cunha@uerj.br](mailto:amico.cunha@uerj.br) (A. Cunha Jr), [rebeccam@colorado.edu](mailto:rebeccam@colorado.edu) (R.E. Morrison).

<https://doi.org/10.1016/j.simpa.2022.100252>

Received 29 December 2021; Received in revised form 9 February 2022; Accepted 9 February 2022

# ARBO

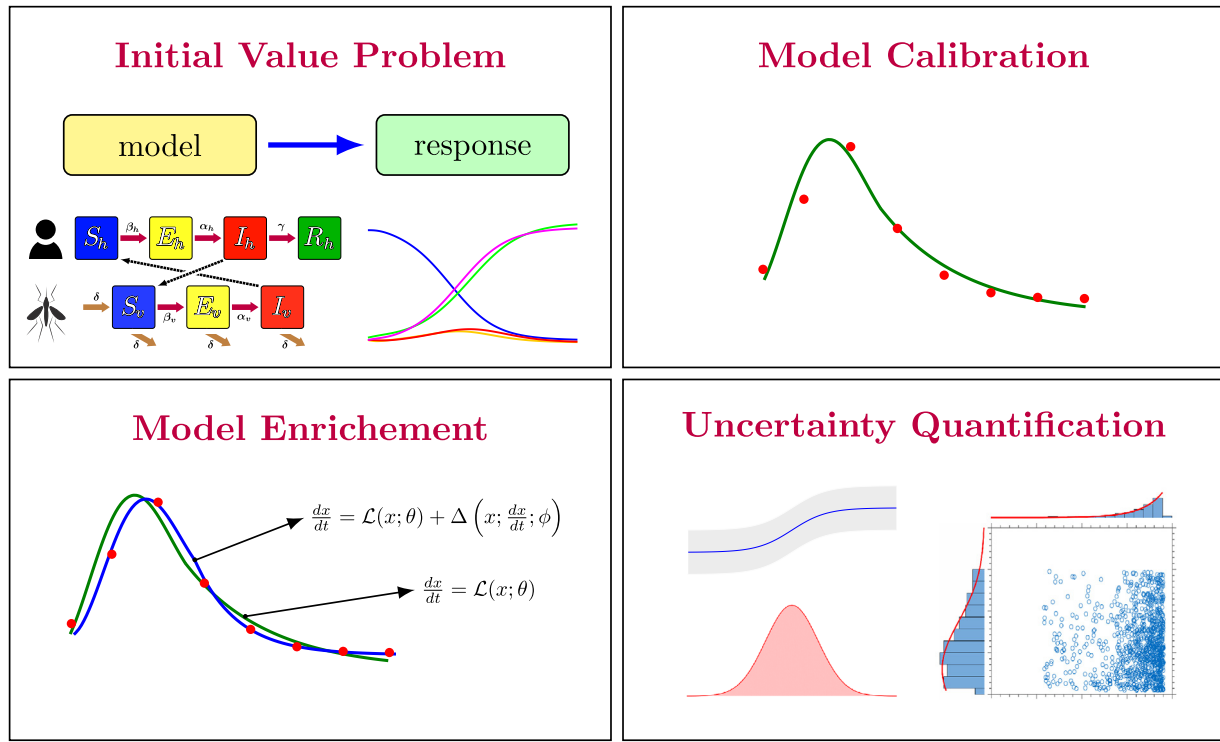


Fig. 1. Schematic representation of ARBO package modules.

## 2. Software details

ARBO provides a toolbox for vector-borne dynamics on a host population, encompassing: (i) an initial value problem (IVP) solver; (ii) an inverse problem module for (deterministic) model calibration, potentially using real data; (iii) a model enrichment module which constructs a discrepancy operator to compensate for epistemic deficiencies (uncertainties) in the structure of the mathematical model; and (iv) an uncertainty quantification module which propagates the aleatory uncertainties from the model parameters and initial conditions through the mathematical model to its response. The package is a didactic and intuitive tool written in Matlab and C++ languages. Each module in the package is briefly described in Fig. 1.

**Initial Value Problem:** In the IVP module, an SEIR-SEI double population compartmental model [9,10] displays the time series evolution for several groups of interest in the host and vector populations. One-population models, like the classic SIR model, could also be implemented and would be compatible with the other modules; however, arbovirus studies require that both populations (human and vectors) be modeled since cross-infection causes most transmission. Additionally, to easily compare with real data, the cumulative number of infectious humans and the number of new cases are shown. The routines allow for simple user control of the time unit, so that the response can match available real data as usually displayed in epidemiological bulletins (days, weeks, months, etc.).

**Model Calibration:** This module solves an inverse problem that takes one model response—number of new cases per unit time—and compares it with experimental data. The best set of parameters (inputs), i.e., the parameter values for which the model response most closely

reproduces the epidemiological data, is found via a nonlinear least squares minimization of the errors. In particular, the optimization uses the Trust-Region-Reflective method [11], which requires initial guesses for the parameters and allows for user-provided minimum and maximum values of each input.

**Model Enrichment:** To address the (epistemic) modeling errors present in the SEIR-SEI compartmental model, an embedded discrepancy operator improves the model response using extra information about the existing state variables (such as state or derivative information). This strategy was first established for chemical kinetics systems [12] and then adapted for epidemiological applications [7]. Note that the quantity of interest is only modified indirectly; that is, the model equations are enriched, which in turn affects any model outputs. This method is in contrast to many discrepancy models which directly adjust the model output to the data, and is consistent with the philosophy that model errors should be corrected via modeling hypotheses.

**Uncertainty Quantification:** This module analyzes, through a set of statistical measures, how parameter and initial condition uncertainties are propagated to the model output [13–15]. The input PDF and possible correlations therein can thus be observed, while the output PDF reveals critical information about the outbreak, such as a specified confidence interval for the total number of cases. Such analysis has significant importance for prediction and quantitative assessments, since it is very difficult to correctly estimate biological values for arbovirus parameters. Furthermore, outbreaks usually start to be documented long after their actual onset, so initial conditions are typically subject to large uncertainty as well.

### 3. Impact overview

The recent worldwide outbreaks of arboviruses increased the demand for relevant applied research on mathematical epidemiology [16]. The ARBO package may help in the quantitative and qualitative simulation of possible scenarios for future outbreaks. Easy-to-use, the code can readily adapt to a wide range of model configurations [6,7,9]. While these studies focused on the Brazilian outbreak, they showed how to construct robust analyzes for a general epidemic under several sources of uncertainties. The current and the above articles will guide users to extend this methodology to future epidemics, whether through real, predicted, or exploratory simulations.

Besides the research contributions, the ARBO package is currently adopted by courses on Uncertainty Quantification and Nonlinear Dynamics at Rio de Janeiro State University, in Brazil. Furthermore, a previous version of this code was used as a baseline structure for the modeling module of EPIDEMIC: Epidemiology Educational Code [17], a pedagogical toolbox for teaching introductory mathematical epidemiology. In this way, the ARBO package has inspired the development of additional tools for mathematical epidemiology research, each tailored to complement the existing background of mathematicians, physicists, engineers, and many other STEM professionals who may be interested in this area.

### 4. Final remarks

ARBO is a computational package for vector-borne dynamics of infectious diseases that provides a start-to-finish analysis of an outbreak model, including simulation, model calibration, model enrichment, and uncertainty quantification. Built in a cohesive and intuitive way, the package can be adapted for a diverse range of two-population compartmental disease models. Observational data can be directly imported for investigation and selection of suitable parameters (inputs). The different modules account for multiple types of uncertainties, as well as errors stemming from model hypotheses, in a comprehensive manner, allowing for robust predictions and analysis over multiple epidemic scenarios. The package has been used successfully in previous studies of real epidemics and also as a didactic tool for students of mathematical epidemiology and nonlinear dynamics.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

The first three authors were supported by 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). The fourth author was supported in part by the Johnson & Johnson Foundation and its Women in STEM2D Scholars Program, USA.

### References

- [1] M.R. Donalizio, A.R.R. Freitas, A.P.B. Von Zuben, Arboviruses emerging in Brazil: challenges for clinic and implications for public health, *Rev. Saúde Públ.* 46, <http://dx.doi.org/10.1590/S1518-8787.2017051006889>.
- [2] Pan American Health Organization, Epidemiological update: Dengue and other arboviruses - 10 June 2020, 2020, <https://www.paho.org/pt/documents/epidemiological-update-dengue-and-other-arboviruses-10-june-2020>. (Accessed 18 June 2021).
- [3] J. Mlakar, et al., Zika virus associated with microcephaly, *N. Engl. J. Med.* 374 (2016) 951–958, <http://dx.doi.org/10.1056/NEJMoa1600651>.
- [4] R. Thompson, J.M. Del Campo, D. Constenla, A review of the economic evidence of aedes-borne arboviruses and aedes-borne arboviral disease prevention and control strategies, *Expert Rev. Vaccines* 19 (2) (2020) 143–162, <http://dx.doi.org/10.1080/14760584.2020.1733419>.
- [5] S. Towers, F. Brauer, C. Castillo-Chavez, A.K.I. Falconar, A. Mubayi, C.M.E. Romero-Vivas, Estimate of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission, *Epidemics* 17 (2016) 50–55, <http://dx.doi.org/10.1016/j.epidem.2016.10.003>.
- [6] E. Dantas, M. Tosin, A. Cunha Jr., An uncertainty quantification framework for a zika virus epidemic model, *J. Comp. Int. Sci.* 10 (2) (2019) 91–96, <http://dx.doi.org/10.6062/jcis.2019.10.02.0163>.
- [7] R. Morrison, A. Cunha Jr., Embedded model discrepancy: A case study of Zika modeling, *Chaos* 30, 051103, <http://dx.doi.org/10.1063/5.0005204>.
- [8] S.T. Ogunlade, A.I. Adekunle, M.T. Meehan, D.P. Rojas, E.S. McBryde, Modeling the potential of wau-wolbachia strain invasion in mosquitoes to control aedes-borne arboviral infections, *Sci. Rep.* 10, (16812), <http://dx.doi.org/10.1038/s41598-020-73819-1>.
- [9] E. Dantas, M. Tosin, A. Cunha Jr, Calibration of a SEIR-SEI epidemic model to describe the zika virus outbreak in Brazil, *Appl. Math. Comput.* 338 (2018) 249–259, <http://dx.doi.org/10.1016/j.amc.2018.06.024>.
- [10] A.J. Kucharski, S. Funk, R.M. Eggo, H.P. Mallet, W.J. Edmunds, E.J. Nilles, Transmission dynamics of zika virus in island populations: a modelling analysis of the 2013–14 french polynesia outbreak, *PLoS Negl. Trop. Dis.* 10 (5) <http://dx.doi.org/10.1371/journal.pntd.0004726>.
- [11] T.F. Coleman, Y. Li, An Interior Trust Region approach for nonlinear minimization subject to bounds, *SIAM J. Optim.* 6 (2) (1996) 418–445, <http://dx.doi.org/10.1137/0806023>.
- [12] R.E. Morrison, T.A. Olive, R.D. Moser, Representing model inadequacy: A stochastic operator approach, *SIAM/ASA J. Uncertain. Quantif.* 6, 457–496, <http://dx.doi.org/10.1137/16M1106419>.
- [13] A. Cunha Jr., Modeling and Quantification of Physical Systems Uncertainties in a Probabilistic Framework, Springer, Cham, 2017, pp. 127–156, [http://dx.doi.org/10.1007/978-3-319-55852-3\\_8](http://dx.doi.org/10.1007/978-3-319-55852-3_8), Probabilistic Prognostics and Health Management of Energy Systems,
- [14] C. Soize, Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering, Springer, Cham, 2017, <http://dx.doi.org/10.1007/978-3-319-54339-0>.
- [15] R. Ghanem, D. Higdon, H. Owahdi, *Handbook of Uncertainty Quantification*, Springer International Publishing, 2017.
- [16] R. Jones, M.A. Kulkarni, T.M.V. Davidson, RADAM-LAC Research Team, B. Talbot, Arbovirus vectors of epidemiological concern in the americas: A scoping review of entomological studies on zika, dengue and chikungunya virus vectors, *PLoS One* 15 (2) (2020) e0220753, <http://dx.doi.org/10.1371/journal.pone.0220753>.
- [17] B. Pavlack, et al., EPIDEMIC - epidemiology educational code, 2021, [www.EpidemicCode.org](http://www.EpidemicCode.org).