Original software publication

epidWaves: A code for fitting multi-wave epidemic models

Americo Cunha Jr [a],[*], Fernando da Conceição Batista [b], Paulo Roberto de Lima Gianfelice [c], Ricardo Sovek Oyarzabal [c], Jose Mario Vicensi Grzybowski [d], Elbert E.N. Macau [c]

[a] Rio de Janeiro State University, Rio de Janeiro, Brazil
[b] Polytechnic Institute of Leiria, Leiria, Portugal
[c] Federal University of São Paulo, São José dos Campos, Brazil
[d] Federal University of Fronteira Sul, Erechim, Brazil

ARTICLE INFO

ABSTRACT

The COVID-19 pandemic has given rise to a great demand for computational models capable of describing and inferring the evolution of an epidemic outbreak in the short term. In this sense, we introduce **epidWaves**, a package that provides a framework for fitting multi-wave epidemic models to data from actual outbreaks of COVID-19 and other infectious diseases.

Code metadata

Current code version	v1.0
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2022-94
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/0219197/tree/v3
Legal Code License	MIT license
Code versioning system used	git
Software code languages, tools, and services used	Matlab and Python
Compilation requirements, operating environments & dependencies	
If available Link to developer documentation/manual	https://github.com/americocunhajr/epidWaves
Support email for questions	americo.cunha@uerj.br

1. Introduction

Using computational models to study past epidemics and make predictions for ongoing outbreaks is not something new or has only become popular in the COVID-19 pandemic. Mathematical tools such as differential equations, statistical regressors, complex networks, etc have been present in the world of epidemiology (biology in a broader sense) for many decades [1–7], being used in the study of diseases such as malaria [8,9], Dengue [10,11], Zika virus and other arboviruses [12–14], etc. Meanwhile, the global health emergency imposed by the recent COVID-19 pandemic has generated numerous original challenges for computational epidemiology [15]. Among them, we can highlight the development of robust data-driven predictive models to be used in

real-time decision making, especially to estimate (in the short term) the expected number of new cases and deaths.

Data-driven models employing phenomenological equations to make predictions or to represent past outbreak data are very useful. Once they do not present excessive complexity or rely on information that is difficult to obtain during the epidemic, as is typically the case with compartmental models based on differential equations, they are very appealing in this context of real-time (or near-real-time) analysis. Therefore, we introduce **epidWaves**, a Matlab/Python package for fitting multi-wave epidemic models to data from COVID-19 and other infectious diseases.
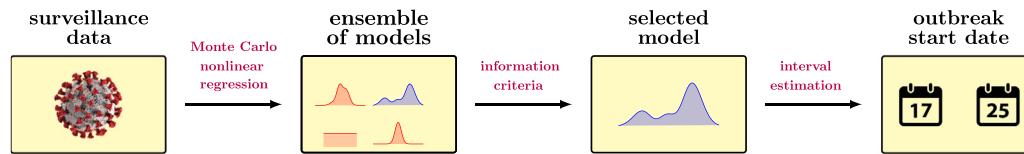
Fig. 1. Schematic representation of **epidWaves** framework.

2. Software details

The package **epidWaves** implements the epidemiological data analysis framework proposed by Gianfelice et al. [16], which is illustrated in Fig. 1. In this methodology, epidemic surveillance data are combined with nonlinear statistical regressors [17,18] and Monte Carlo simulations [19,20] to generate a family of predictive models for the evolution of reported cases and deaths associated with an epidemic outbreak. The available models correspond to logistic curves with multiple modes, describing epidemic dynamics with several waves of contagion (each with three phases: expansion, transition, and exhaustion of the outbreak). The user selects the desired amount of waves or can test different models for the same data. The model which presents the best compromise between good adherence to data and simplicity is chosen based on statistical information criteria. Such a selected model can predict reported cases and deaths during the outbreak in a short-term time horizon or extract various information about previous outbreaks. For instance, with such a model, it is possible to estimate the start date of each wave of contagion (as done by Gianfelice et al. [16]). Alternatively, one can infer other quantitative information, such as the "velocity" of expansion of a contagion wave (through the infection rate), the date of a wave peak (which characterizes the beginning of the outbreak exhaustion phase), or even an estimated date for the beginning of a contagion wave. Such metrics may allow an epidemiologist to assess the severity of a wave of contagion, compare different waves, and investigate the possible factors that triggered the wave. Matlab and Python codes for each step of the framework are available in the package.

3. Impact overview

Epidemics are recurrent in the history of humanity, and due to globalization, they tend to be increasingly frequent and challenging. Making real-time (or near-real-time) decisions driven by computer models in such scenarios has been critical in the COVID-19 pandemic and will be in future large-scale epidemics. The **epidWaves** package is a valuable tool for such a task, which due to its simplicity, can be used by professionals with relatively modest mathematical training. In addition to building predictive models, the statistical regressors obtained with this package can provide qualitative and quantitative descriptions of past outbreak data, improve understanding of the epidemic dynamics, and extract critical information (e.g., the start date of a contagion wave). Examples of application in this sense can be seen in the Refs. [16,21]. In the first work, the authors use the **epidWaves** framework to estimate the starting date of COVID-19 contagion waves in Rio de Janeiro city, between 2020 and 2021. In the second, the authors investigate the dynamics of COVID-19 in Portugal, investigating details of the 2020 outbreak. In addition, it is worth mentioning that the independent initiative *COVID-19: Observatório Fluminense*[1] used initial versions of this code during its epidemic surveillance work [22].

---

[1] Interdisciplinary group of researchers that monitored the advance of COVID-19 in Brazil in the first two years of the pandemic. Further details can be seen in **www.covid19rj.org**

Due to the versatility and simplicity of the package, **epidWaves** also has a tremendous educational appeal. It is also worth mentioning that the epidemic curve fitting methodology employed here and early versions of the underlying Matlab code served as a basis for the forecasting module of the educational code EPIDEMIC [23]. Thus, this novel package joins other tools for teaching computational epidemiology, such as EPIDEMIC [23] and ARBO [24], both developed by researchers from Rio de Janeiro State University and collaborators.

4. Final remarks

epidWaves is a simple and powerful tool to fit statistical multi-wave models to epidemic data from COVID-19 and other infectious diseases. The analyzes performed with the statistical models built with this package can be carried out during the epidemic outbreak or a posteriori, with a view to better understanding the evolution of the disease in the latter case and guiding immediate decision-making in the first case. The code also has enormous potential as an educational tool for computational epidemiology, already being used for both research and teaching activities.

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.

References

[1] H.W. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (4) (2000) 599–653, http://dx.doi.org/10.1137/S0036144500371907.

[2] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200–3203, http://dx.doi.org/10.1103/PhysRevLett.86.3200.

[3] F. Brauer, C. Castillo-Chávez, Mathematical Modells in Population Biology and Epidemiology, Springer, 2001.

[4] F. Brauer, P. van den Driessche, J. Wu (Eds.), Mathematical Epidemiology, Springer, 2008.

[5] R. Morrison, A. Cunha Jr, Embedded model discrepancy: A case study of Zika modeling, Chaos 30 (051103) (2020) http://dx.doi.org/10.1063/5.0005204.

[6] M. Tosin, A.M.A. Côrtes, A. Cunha, A tutorial on sobol' global sensitivity analysis applied to biological models, in: F.A.B. Silva, N. Carels, M.T. Santos, F.J.P. Lopes (Eds.), Networks in Systems Biology: Applications for Disease Modeling, Springer International Publishing, Cham, 2020, pp. 93–118, http://dx.doi.org/10.1007/978-3-030-51862-2_6.

[7] G.S. Costa, W. Cota, S.C. Ferreira, Outbreak diversity in epidemic waves propagating through distinct geographical scales, Phys. Rev. Res. 2 (2020) 043306, http://dx.doi.org/10.1103/PhysRevResearch.2.043306.

[8] S. Mandal, R.R. Sarkar, S. Sinha, Mathematical models of malaria - A review, Malar. J. 10 (1) (2011) 202, http://dx.doi.org/10.1186/1475-2875-10-202.

[9] M.F.C. Gomes, C.T. Codeço, L.S. Bastos, R.M. Lana, Measuring the contribution of human mobility to malaria persistence, Malar. J. 19 (1) (2020) 404, http://dx.doi.org/10.1186/s12936-020-03474-4.

[10] C.T. Codeço, D.A. Villela, F.C. Coelho, Estimating the effective reproduction number of dengue considering temperature-dependent generation intervals, Epidemics 25 (2018) 101–111, http://dx.doi.org/10.1016/j.epidem.2018.05.011.

[11] P.M. Luz, B.V.M. Mendes, C.T. Codeço, C.J. Struchiner, A.P. Galvani, Time series analysis of dengue incidence in Rio de Janeiro, Brazil, Am. J. Trop. Med. Hyg. 79 (6) (2008) 933–939, http://dx.doi.org/10.4269/ajtmh.2008.79.933.

[12] 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) (2016) http://dx.doi.org/10.1371/journal.pntd.0004726.

[13] 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.

[14] E. Dantas, M. Tosin, A. Cunha Jr, An uncertainty quantification framework for a Zika virus epidemic model, J. Comput. Interdiscipl. Sci. 10 (2019) 91.

[15] E. Kuhl, Computational Epidemiology: Data-Driven Modeling of COVID-19, Springer, 2021.

[16] P.R.L. Gianfelice, R.S. Oyarzabal, A. Cunha, J.M.V. Grzybowski, F.C. Batista, E.E.N. Macau, The starting dates of COVID-19 multiple waves, Chaos 32 (3) (2022) 031101, http://dx.doi.org/10.1063/5.0079904.

[17] L. Wasserman, All of Statistics, Springer, 2004.

[18] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, second ed., Springer, 2009.

[19] D.P. Kroese, T. Taimre, Z.I. Botev, Handbook of Monte Carlo Methods, Wiley, 2011.

[20] A. Cunha Jr, R. Nasser, R. Sampaio, H. Lopes, K. Breitman, Uncertainty quantification through Monte Carlo method in a cloud computing setting, Comput. Phys. Comm. 185 (2014) 1355–1363, http://dx.doi.org/10.1016/j.cpc.2014.01.006.

[21] F.C. Batista, A. Cunha Jr, Inferência dos estágios iniciais da COVID-19 em Portugal, in: XXIII Congresso Brasileiro de Automática, CBA 2020, Santa Maria, Brazil, 2020.

[22] A. Cunha Jr, et al., Relatório 01 progresso da COVID-19 no Brasil e no estado do Rio de Janeiro 21ª semana epidemiológica do calendário 2020 (17/5/2020 até 23/5/2020), 2020, http://dx.doi.org/10.12957/eduerj.covid19rj.relatorio1, COVID-19: Observatório Fluminense (in Portuguese).

[23] B. Pavlack, M. Grave, E. Dantas, J. Basilio, L. de la Roca, J.a. Norenberg, M. Tosin, L. Chaves, D. Matos, M. Issa, R. Luo, A. Guyt, L. Soares, R. Burgos, L. Lovisolo, A. Cunha, EPIDEMIC - Epidemiology educational code, J. Open Source Educ. 5 (2022) 149, http://dx.doi.org/10.21105/jose.00149.

[24] M. Tosin, E. Dantas, A. Cunha Jr, R.E. Morrison, ARBO: Arbovirus modeling and uncertainty quantification toolbox, Software Impacts 12 (2022) 100252, http://dx.doi.org/10.1016/j.simpa.2022.100252.