The Journal of Open Source Education

EPIDEMIC: Epidemiology Educational Code

Bruna Pavlack[1,2], **Malú Grave**[3], **Eber Dantas**[3], **Julio Basilio**[4], **Leonardo de la Roca**[4], **João Pedro Norenberg**[2], **Michel Tosin**[4], **Lucas Chaves**[5], **Diego Matos**[4], **Marcos Issa**[4], **Roberto Luo**[4], **Amanda Cunha Guyt**[6], **Luthiana Soares**[7], **Rodrigo Burgos**[4], **Lisandro Lovisolo**[4], **and Americo Cunha Jr**[*4]

1 Federal Institute of Mato Grosso do Sul **2** State University of São Paulo **3** Federal University of Rio de Janeiro **4** Rio de Janeiro State University **5** Federal University of Uberlândia **6** City College of San Francisco **7** Federal University of Pampa

DOI: 10.21105/jose.00149

Software

- Review
- Repository
- Archive

Submitted: 02 March 2021
Published: 05 April 2022

Computational epidemiology

The aim of epidemiology is to understand the causes of a disease in order to predict its evolution and then make decisions to contain the epidemic. Thus, epidemiology is of great importance to society.

The study of a disease spreading leads to the generation of large data sets that requires the use of computational methods to generate epidemiological models and analyze results for decision-making (Tameru et al., 2012). In this context, computational epidemiology arises, which over time has become increasingly multidisciplinary (techniques of epidemiology, biology, mathematics, theoretical computer science, machine learning, etc.) and led to the development of new computational methods to understand and control the spread of disease (Marathe & Ramakrishnan, 2013).

Statement of need

Due to the Covid-19 pandemic, several researchers from different Brazilian institutions organized an initiative called *COVID-19: Observatório Fluminense* (COVID-19RJ) (*COVID-19*, 2020), which aims to monitor the pandemic progress in Brazil, make reliable predictions about the short term evolution and provide high-quality educational material about the mathematical modeling and analysis of COVID-19. To support the research carried out in the COVID-19RJ project, the EPIDEMIC was developed, because through this code it is possible to generate monitoring graphics, trends and disease forecasts. The EPIDEMIC is cited in the reports of the COVID-19RJ project.

In the context of the COVID-19 pandemic, it was observed that computational epidemiology gained prominence, attracting the interest of many people from different areas of knowledge. There are excellent codes available for conducting epidemiological simulations, but these are customized for researchers in the area (Abdulrahman, 2020; Adhikari et al., 2020; Dantas et al., 2018; Hladish et al., 2012; Morrison & Cunha Jr, 2020), so members of COVID-19RJ felt the need to organize EPIDEMIC in a pedagogical way, to collaborate in the training of new researchers. Thus, in addition to EPIDEMIC being a research tool, it is an easy-to-use code that provides a detailed tutorial with several

[*] Corresponding author (americo.cunha@uerj.br).

examples, facilitating the insertion of new researchers in the field of epidemiology and also assisting in teaching courses on computational modeling and epidemiology.

In a mathematical biology course, for example, students are exposed to the derivation of equations from various compartmental models based on nonlinear differential equations, so that the vast majority have no analytical solution. In this context, there is a natural need to use numerical methods to obtain the answer from the mathematical model. Packages like GNU Octave and Matlab have very robust ordinary differential equations (ODEs) solvers that make this task much easier. The EPIDEMIC tutorials provide pedagogical activities in this sense, as they illustrate, via relatively streamlined and very well organized codes, how to perform the numerical integration of these models based on ODEs. The programming structure followed in the tutorial, with several compartmental models, also aims to show students how easy it is to carry out such numerical integration for another type of compartmental model that the student can find in the literature or develop himself/herself.

The benefits of the EPIDEMIC tutorial go beyond the simple exercise of programming numerical integration, the analysis of the response curves of different models allows students to develop intuition about the dynamic behavior of dynamic epidemiological systems. For example, when seeing in the same figure the 3 curves of the SIR model (susceptible, infected and recovered), the student can see that the initial growth of an outbreak is accompanied at the same time by a significant reduction in the number of susceptibles, being succeeded by an increase in the number of retrieved. An interesting analogy in this case that aids in the students' understanding is the following: be the infected as the intensity of a fire, the susceptible as the available firewood, and the recovered as the burnt firewood after a fire. At the beginning there is a lot of firewood available, a large fire ensues, then, as the firewood burns, the intensity of the fire decreases.

The EPIDEMIC codes have already been used in a course on numerical and computational methods in Rio de Janeiro State University to present the part of polynomial regression and exponential curve fitting. In this course, the code was used on three fronts, according to the EPIDEMIC modules: modeling, trends and forecast. In the module modeling, the compartmental models were presented as computer simulators, students run the codes as a black box to familiarize themselves with simulation basics, and get acquainted with the Octave software. The code is also used in interpolation classes. We want to obtain a polynomial function that describes the number of infected people at the peak of the epidemic as a function of the transmission rate beta. Students simulate some cases by varying the beta parameter, obtaining some values for the number of infected at peak. Then they interpolate polynomial curves to describe this functional relationship. Students are invited to reflect which polynomial is most representative, that is, which makes the most sense. The module trends was used in the data visualization class, to train students on how to show different information in a clear, objective, effective and graphically attractive way. In this exercise, the effect of normalizing was also shown, to remove the scaling effect of population size (eg infected vs time / infected by 1M inhab vs time, when normalizing the curves approach the same level, without local normalization of larger population has much larger numbers, which may not be true when normalizing). And, the module forecast was used to train the students in the regression part (curve fitting), they used the COVID-19 epidemic data as observations, and looked for polynomial and exponential curves that fit the start of the outbreak (exponential phase). Bellow it is presented a brief description of the code and modules that make up the EPIDEMIC.

EPIDEMIC code

The EPIDEMIC code is a suite of basic software for epidemiology that is suitable for analyzing indicators of an epidemic evolution as well as to construct basic compartmental models for qualitative and quantitative analysis. It is developed in an easy to use style, with very lean and well-documented codes. The package also includes an instructional tutorial that gives the end user information about the type of analysis and forecasts that can be obtained with the suite, as well as an overview of the limitations of each model available in the code. The EPIDEMIC code is a didactic and intuitive pedagogical tool for audiences interested in mathematical epidemiology. The EPIDEMIC is developed on the free software platform GNU Octave and is available on a website[1], where one can find a GitHub repository link that directs to the suite source code. The EPIDEMIC code is also compatible with MATLAB proprietary software. The fact that EPIDEMIC is an open package, allows the user to better understand the mathematical structure of the model and is accessible to a large audience. It is interesting that the user has previous knowledge of GNU Octave or MATLAB, but as the organization of the EPIDEMIC code is done in a very didactic way, those students who have a first access to these programming languages will also be able to carry out the simulations. The figure 1 shows the EPIDEMIC logo and the three EPIDEMIC modules: modeling, trends and forecasts.

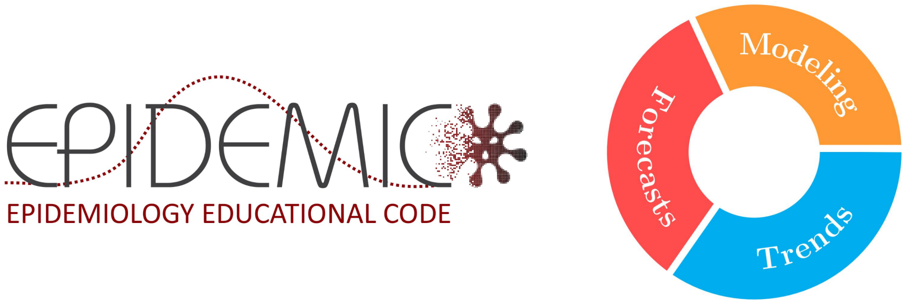

Figure 1: Illustration of the EPIDEMIC code logo and EPIDEMIC modules.

Modeling

In the modeling module, compartmental models, described by differential equations, are used to simulate population dynamics during an epidemic. The compartmental models available are: SIR, SEIR, SIRG, SEIRD, and SEIAHRD.

It is possible with EPIDEMIC to plot the curves of the aforementioned compartmental models and thus analyze the epidemiological dynamics. To do this, just change the values of the parameters of the epidemic that you want to analyze in the code. It is also possible to carry out analyzes, for example, of $\mathcal{R}_0$ in relation to β, or else of the number of deaths in relation to β. In Figure 2 is possible to see an example time series graph of the SEIAHRD compartmental model.

[1] www.EpidemicCode.org

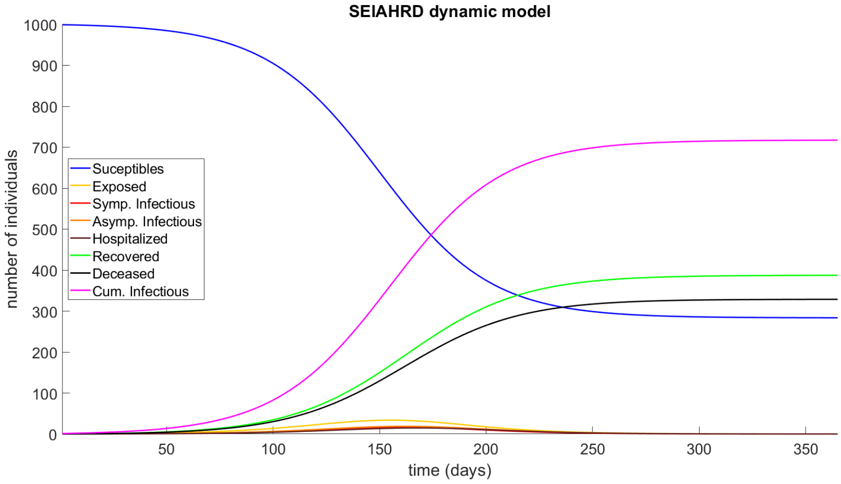

Figure 2: Time series graph of the SEIAHRD dynamic model generated in the modeling module of EPIDEMIC.

Trends

In the trends module, it is possible to monitor, through graphic resources, the behavior of epidemics in countries, states or cities. The analyses are performed using two basic visualization strategies: that of contagion and mortality and the progress of the epidemic. The types of graphs that can be generated in the trends module are: accumulated deaths, accumulated cases, death progress, case progress, weekly deaths, weekly cases, mortality, prevalence, weekly deaths, and incidence. The Figure 3 shows an example of a trend graph generated in EPIDEMIC. This graph shows the total number of deaths per 1M inhabitants evolution of the COVID-19 pandemic in several countries.

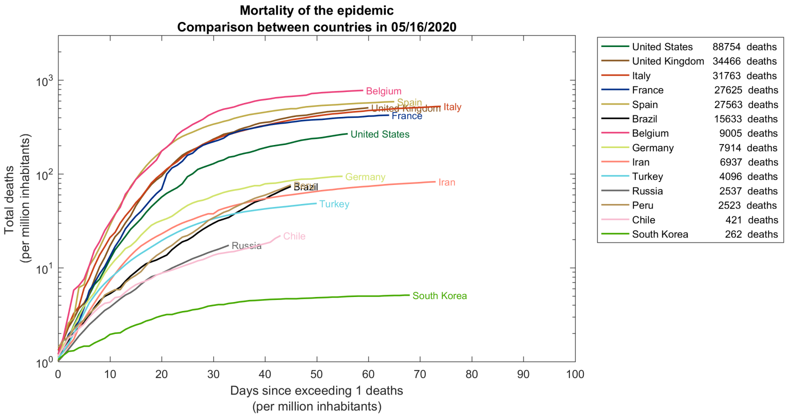

Figure 3: Trend graph of total number of deaths per 1M inhabitants evolution of the COVID-19 pandemic in several countries generated in EPIDEMIC.

Forecasts

In the forecasts module, a statistical regressor is used to obtain forecasts about the short term behavior of the epidemic curves. The method used to obtain the regressor is the classic *Ordinary Least Squares*. In preparing the forecasting code, the following points were taken into account:

- Consider the last five days of the data sample, as these reflect the most recent trend;
- Insert the last five days on the logarithmic scale, as it facilitates visualization in case of exponential growth;
- Plot the predicted values within the estimated reliability envelope. Confidence band is 95%.

The Figure 4 shows a forecast graph of the total cases of COVID-19 in Brazil generated at EPIDEMIC.

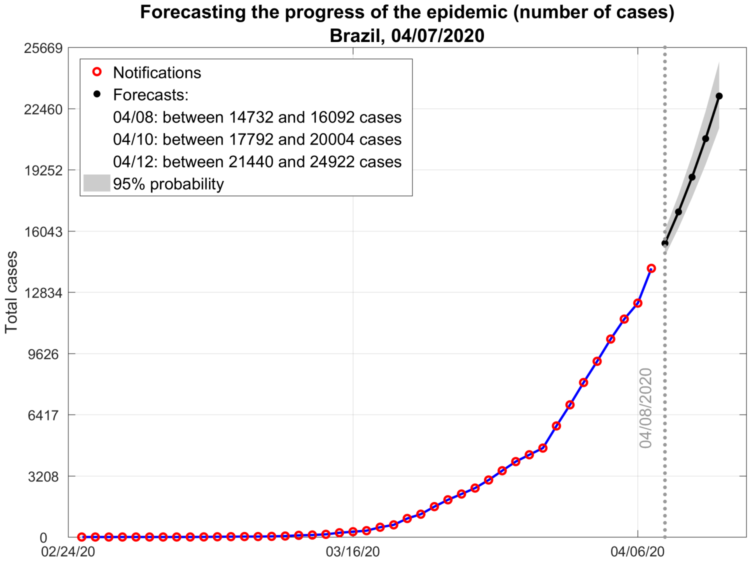

Figure 4: Forecast graph of the total cases of COVID-19 in Brazil generated at EPIDEMIC. The gray color shows the 95% confidence band.

Implementation

For the epidemic implementation, in the modeling module, the package includes the files *main_X.m* and *rhs_X.m*. The ”X” in the name represents the compartmental model (SIR, SEIR, SIRG, SEIRD, and SEIAHRD). The file *main* defines the parameters and calculates the reproduction numbers and plots the results of the time series. The File *rhs* defines the ODE system used by the main file. In the trends module, there is the file *epidemic_trends.m*, which is the main file to generate graphs on the numbers of cases and deaths by epidemic in the countries of interest. And in the forecasts module, there is the file *epidemic_forecasts.m*, which is the main file to generate the forecast graphs of accumulated cases and accumulated deaths from an epidemic.

To check the restrictions on the use of EPIDEMIC routines, the ”test” folder brings together a set of ”verification scripts” to individually cover each possible error in these routines. To facilitate its use, EPIDEMIC has a tutorial, in English and Portuguese, with examples and explanations about the code. The codes and the EPIDEMIC tutorial are constantly being updated, according to the needs that come up.

This educational code proves to be an important didactic tool for epidemiological analysis, as it is available in a transparent, accessible and reproducible way (Chatterjee, 2020). Therefore, it is also an important tool for the development of research.

References

Abdulrahman, I. (2020). *SimCOVID: Open-source simulation programs for the COVID-19 outbreak*. https://doi.org/10.1101/2020.04.13.20063354

Adhikari, R., Bolitho, A., Caballero, F., Cates, M. E., Dolezal, J., Ekeh, T., Guioth, J., Jack, R. L., Kappler, J., Kikuchi, L., Kobayashi, H., Li, Y. I., Peterson, J. D., Pietzonka, P., Remez, B., Rohrbach, P. B., Singh, R., & Turk, G. (2020). *Inference, prediction and optimization of non-pharmaceutical interventions using compartment models: The PyRoss library*. https://arxiv.org/abs/2005.09625

Chatterjee, N. (2020). *Transparency, reproducibility, and validity of COVID-19 projection models*. https://towardsdatascience.com/transparency-reproducibility-and-validity-of-covid-19

COVID-19: Observatório Fluminense. (2020). https://www.covid19rj.org

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

Hladish, T., Melamud, E., Barrera, L., Galvani, A., & Meyers, L. (2012). EpiFire: An open source C++ library and application for contact network epidemiology. *BMC Bioinformatics*, *13*, 76. https://doi.org/10.1186/1471-2105-13-76

Marathe, M. V., & Ramakrishnan, N. (2013). Recent advances in computational epidemiology. *IEEE Intelligent Systems*, *28*, 96–101. https://doi.org/10.1109/MIS.2013.114

Morrison, R. E., & Cunha Jr, A. (2020). Embedded model discrepancy: A case study of Zika modeling. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *30*, 051103. https://doi.org/10.1063/5.0005204

Tameru, B., Nganwa, D., Bogale, A., Robnett, V., & Habtemariam, T. (2012). The role of computational epidemiology and risk analysis in the fight against HIV/AIDS. *Journal of AIDS & Clinical Research*, *36*. https://doi.org/10.4172/2155-6113.1000e107