

EPIDEMIC: a didactic tool for teaching mathematical epidemiology

Bruna Pavlack^{1,2}, Malú Grave³, Eber Dantas³, Julio Basilio⁴, Leonardo de la Roca⁴, João Pedro Norenberg², Michel Tosin⁴, Lucas Chaves⁵, Diego Matos⁴, Marcos Issa⁴, Roberto Luo⁴, Amanda Cunha Guyt, Luthiana Soares, Rodrigo Burgos⁴, Lisandro Lovisoló⁴, Americo Cunha Jr⁴

¹Federal Institute of Mato Grosso do Sul, Três Lagoas, MS

²State University of São Paulo, Ilha Solteira, SP

³Federal University of Rio de Janeiro, Rio de Janeiro, RJ

⁴Rio de Janeiro State University, Rio de Janeiro, RJ

⁵Federal University of Uberlândia, Uberlândia, MG

Due to the COVID-19 pandemic, there was a growing interest in society in the area of epidemiology. In this context, emerged the initiative *COVID-19: Observatório Fluminense* (COVID19RJ), which is managed by a group of independent researchers affiliated with different Brazilian institutions. COVID19RJ publishes trend and monitoring graphs of the COVID-19 pandemic in Brazil and some countries around the world. Due to this demand, the need for an educational code for research in epidemiology arose, with the objective of encouraging and allowing the entry of more researchers in this area. So, EPIDEMIC was developed, this being a code that is organized in a didactic way and divided into three modules: modeling, trends and forecasts. In the modeling module, compartmental models are used to simulate population dynamics during an epidemic. In the trends module, it is possible to monitor the behavior of epidemics in countries, states or municipalities. And in the forecasts module, a statistical regressor is used to obtain predictions about the short-term behavior of epidemic curves. EPIDEMIC provides a tutorial with explanations and examples of use. The code was developed on the free software GNU Octave and is compatible with the proprietary software MATLAB. EPIDEMIC presents itself as a good aid tool for the teaching and learning process of epidemiology and, consequently, for the increase of researchers in the area.

Palavras-chave. Epidemiology, Educational Code, Compartmental Models, Octave, Trend and Forecast Graphs

1 Introduction

An epidemic arises in a community or region when cases of illness or other health-related events occur beyond normal expectations. For many years, there have been reports of several epidemics that have hit society and with the emergence of new diseases, such as HIV/AIDS and hepatitis C, interest in infectious diseases has increased a lot in recent years. The science that studies epidemics and aims to analyze the distribution of phenomena related to diseases and their conditioning and determining factors in populations is called 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. For epidemiology to reach its objective, it is necessary to obtain and analyze disease data, requiring the use of

¹bruna.pavlack@ifms.edu.br

mathematical tools. Daniel Bernoulli's [3] work on smallpox is generally described as the first model in mathematical epidemiology. In [7] it was proposed that the spread of a communicable disease depends on the number of susceptible and infected individuals, this being the basic idea for the study of compartmental models in epidemiology. In [14] it was developed a simple compartmental model on the dynamics of malaria transmission between mosquitoes and humans, and showed that to contain the spread a reduction in the mosquito population below a critical level would be sufficient. This was the introduction of the basic reproduction number (\mathcal{R}_0) concept. In Kermack and McKendrick [9–11] it was assumed that the compartments are distributed exponentially. It is known that at the beginning of a disease outbreak, there is a very small number of infectious individuals and the transmission of the infection is a stochastic event depending on the pattern of contacts between the population, making it necessary to consider stochastic epidemiological models.

The study of the spread of a disease leads to the generation of large data sets that require the use of computational methods to generate epidemiological models and analyze results for decision making. 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 [12].

The intensity of COVID-19 pandemic in Brazil is above the average compared to other countries, generating several demands in terms of monitoring and forecasting the epidemic's progress on Brazilian soil. In this context, several researchers from different Brazilian institutions organized an initiative called COVID-19: Observatório Fluminense (COVID19RJ) [5], 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 COVID19RJ project, the EPIDEMIC was developed, because through this code it is possible to generate monitoring graphics, trends and disease forecasts. 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 [1, 2, 6, 8, 13], so members of COVID19RJ 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 tool that provides a detailed tutorial with several examples, facilitating the insertion of new researchers in the field of epidemiology.

2 EPIDEMIC Code

The EPIDEMIC code (Figure 1) is a suite of basic software for research in epidemiology. The code is suitable for analyzing indicators of the evolution of an epidemic, as well as for building basic compartment models for qualitative and quantitative analyzes. The package also includes an instructional tutorial that provides the user with information about the type of analysis and predictions that can be obtained with the package, as well as an overview of the limitations of each model available in the code. The EPIDEMIC code is organized in an intuitive way and with the objective of providing a didactic tool that helps in the teaching and learning process in the field of epidemiology. The EPIDEMIC code is developed on the free software platform GNU Octave and is available on the site ², where you can find a link to the GitHub repository that directs to the source code of the suite. The code is also compatible with proprietary software MATLAB. The

²www.EpidemicCode.org

EPIDEMIC code structure consists of three modules: modeling, trends and predictions. The next subsections cover topics about the modules that make up the EPIDEMIC and its implementation.



Figure 1: Illustration of the EPIDEMIC code logo.

2.1 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. The SIR and SEIR models are simpler and more widely known. On GitHub it is possible to find several codes that include the SIR and SEIR models, but none of them are organized in modules (modeling, trends and forecasts) and have an educational sense. Besides that, the EPIDEMIC presents examples of more complex compartmental models.

In the SIR model, the compartments considered are: susceptible, infected and recovered. In Figure 2 it is possible to observe a schematic representation of the SIR model, where β is transmission rate and γ is recovery rate. The basic reproduction number is $\mathcal{R}_0 = \frac{\beta}{\gamma}$.



Figure 2: Schematic representation of the SIR model.

The differential equations, as a function of time, for the dynamics of the SIR model are:

$$\begin{aligned}
 \frac{dS}{dt} &= -\beta S \frac{I}{N} && \text{(rate of susceptible)} \\
 \frac{dI}{dt} &= \beta S \frac{I}{N} - \gamma I && \text{(rate of infected)} \\
 \frac{dR}{dt} &= \gamma I && \text{(rate of recovered)}
 \end{aligned} \tag{1}$$

In the SEIR model, the compartments considered are: susceptible, exposed, infected and recovered. The SEIRD model has the addition of the deceased compartment. In the SEIAHRD model, the compartments that are considered to be more than those already mentioned are: asymptomatic, infectious and hospitalized. The SIRG model has compartmental classes susceptible, infected, recovered and guarded. In this model, the guarded compartment is in the sense of confined.

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 3 is possible to see an example time series graph of the SIRG compartmental model.

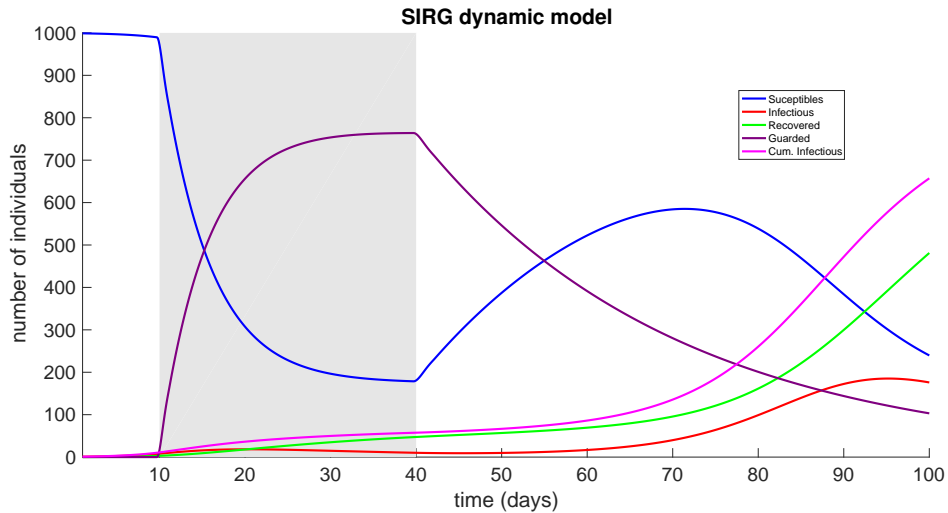


Figure 3: Time series graph of the SIRG dynamic model generated in the modeling module of EPIDEMIC. The gray band on the graph represents the quarantine period of the guarded. In this case, the quarantine starts on the 10th day and ends on the 40th day.

2.2 Trends

In the trends module, it is possible to monitor, through graphic resources, the behavior of epidemics in countries, states or cities. The analyzes are performed using two basic visualization strategies: that of contagion and mortality and the progress of the epidemic. In Table 1 it is possible to observe the types of graphs that can be generated in the trends module.

Graph type	Presented data
Accumulated deaths	absolute value (total deaths per time)
Accumulated cases	absolute value (total cases per time)
Death progress	absolute value (new deaths per total deaths)
Case progress	absolute value (new cases per total cases)
Weekly deaths	absolute value (new deaths per time)
Weekly cases	absolute value (new cases per time)
Mortality	per million inhabitants (total deaths per million)
Prevalence	per million inhabitants (total cases per million over time)
Weekly deaths	per million inhabitants (new weekly deaths per million)
Incidence	per million inhabitants (new weekly cases per million)

Table 1: Types of graphics generated by EPIDEMIC in the trends module.

The Figure 4 shows an example of a trend graph generated in EPIDEMIC. This graph analyzes the contagion evolution of the epidemic, showing the total number of cases as a function of time of the COVID-19 pandemic in several countries. The same analysis can be applied to the number of deaths (evolution of mortality).

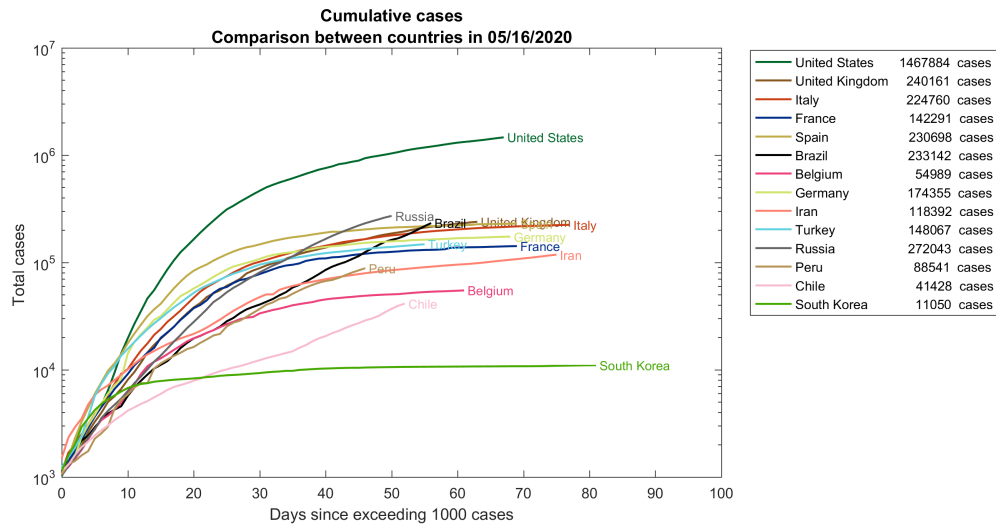


Figure 4: Graph of total number of cases in relation to time of the COVID-19 pandemic in several countries.

2.3 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 5 shows a forecast graph of the total deaths of COVID-19 in Brazil generated at EPIDEMIC.

2.4 Implementation

For the epidemic implementation, the package includes the files that are described in Table 2. In the name of the files in the modeling module, the "X" represents the compartmental model (SIR, SEIR, SIRG, SEIRD and SEIAHRD). 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.

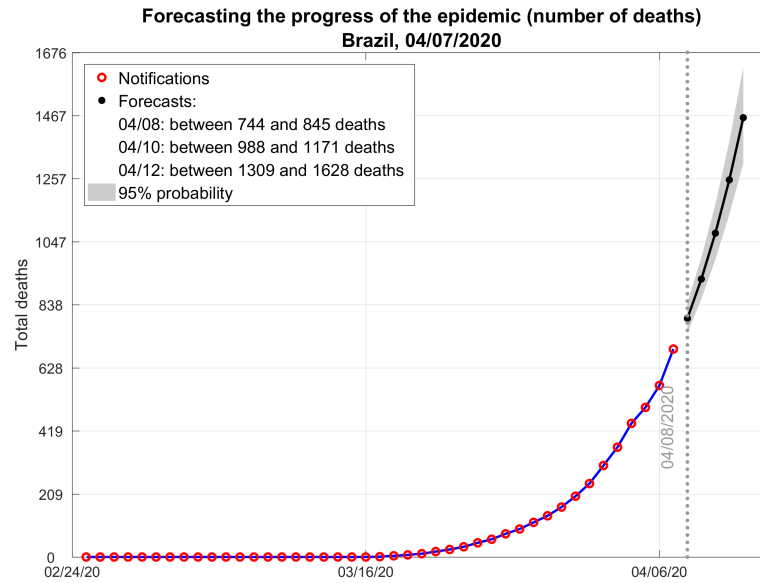


Figure 5: Graph of forecast of total deaths from COVID-19 in Brazil (time-dependent). The gray color shows the 95 % confidence band.

Module	File	Description
Modeling	main_X.m	Main file. Defines the parameters and calculates the reproduction numbers and plots the results of the time series.
	rhs_X.m	Defines the ODE system used by the main file.
Trends	epidemic_trends.m	Main file to generate graphs on the numbers of cases and deaths by epidemic in the countries of interest.
Forecasts	epidemic_forecasts.m	Main file to generate the forecast graphs of accumulated cases and accumulated deaths from an epidemic.

Table 2: EPIDEMIC package files.

3 Conclusion

The EPIDEMIC is an important didactic tool for epidemiological analysis, as it is available in a transparent, accessible and reproducible way [4]. Its organization by modules and the structure of the code facilitates the user’s understanding of the results of epidemiological analysis. In addition, a tutorial is available, in English and Portuguese, with several examples of use and explanations of terms in epidemiology. In this way, EPIDEMIC is also an important tool for the development of research, as it contributes to the insertion of new researchers in the area, in addition to the code allowing the generation of research results in epidemiology.

COVID19RJ uses EPIDEMIC to generate its monitoring and forecasts graphs for the COVID-19 pandemic in Brazil and in some countries. The code is quoted in the reports made available by the group and which can be found on the website www.covid19rj.org.

References

- [1] Abdulrahman, I. SimCOVID: Open-Source Simulation Programs for the COVID-19 Outbreak, 2020. DOI: 10.1101/2020.04.13.20063354.
- [2] 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. and Turk, G. Inference, prediction and optimization of non-pharmaceutical interventions using compartment models: the PyRoss library, 2020.
- [3] Brauer, F. Mathematical epidemiology: Past, present, and future, *Infectious Disease Modelling*, 2, 2017. DOI:10.1016/j.idm.2017.02.001.
- [4] Chatterjee, N. Transparency, Reproducibility, and Validity of COVID-19 Projection Models, Towards Data Science, 2020. Available in <<https://towardsdatascience.com/transparency-reproducibility-and-validity-of-covid-19-projection-models-78592e029f28>>. Access in June 30, 2020.
- [5] COVID-19: Observatório Fluminense, 2020. Available in <<https://www.covid19rj.org>>. Access in March 4, 2020.
- [6] Dantas, E., Tosin, M. and 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.
- [7] Hamer, W. H., Cantab, M. D. and Lond, F. R. C. P. The Milroy Lectures ON EPIDEMIC DISEASE IN ENGLAND — THE EVIDENCE OF VARIABILITY AND OF PERSISTENCY OF TYPE, *The Lancet*, 167:569–574, 1906. DOI:10.1016/S0140-6736(01)80187-2.
- [8] Hladish, T., Melamud, E., Barrera, L., Galvani, A. and Meyers, L. EpiFire: An open source C++ library and application for contact network epidemiology, *BMC Bioinformatics*, 13, 2012. DOI:10.1186/1471-2105-13-76.
- [9] Kermack, W. O. and McKendrick, A. G. A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London*, 115:700–721, 1927. DOI:10.1098/rspa.1927.0118.
- [10] Kermack, W. O. and McKendrick, A. G. Contributions to the mathematical theory of epidemics. II. — The problem of endemicity, *Proceedings of the Royal Society of London*, 138:55–83, 1932. DOI:10.1098/rspa.1932.0171.
- [11] Kermack, W. O. and McKendrick, A. G. Contributions to the mathematical theory of epidemics. III. — Further studies of the problem of endemicity, *Proceedings of the Royal Society of London*, 141:94–122, 1933. DOI:10.1098/rspa.1932.0171.
- [12] Marathe, M. V. and Ramakrishnan, N. Recent Advances in Computational Epidemiology, *IEEE Intelligent Systems*, 28:96–101, 2013, (to appear).
- [13] Morrison, R. E. and Cunha Jr, A. Embedded model discrepancy: A case study of Zika modeling, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30, 2020. DOI:10.1063/5.0005204.
- [14] Ross, R. A. The prevention of malaria (with Addendum), *London: John Murray*, 1911, (to appear).