Study of the Temporal Propagation of Arboviruses in the Region of Recife-PE: Analysis of Climatic Influence using the SIR Model and Recurrent Neural

Guilherme Lopes de Campos[1], Rita de Cássia de Vasconcelos Pedrosa[2 3], Cláudio Tadeu Cristino[4]

[1]*Departamento de Engenharia Química, Universidade Federal Fluminense*
Rua Passos da Pátria, 24210-240, Niterói/Rio de Janeiro, Brasil
guilherme_lc@id.uff.br
[2]*Professora bolsista de produtividade, Centro Universitário Estácio Recife*
Av. Eng. Abdias de Carvalho, 1678, 50720-225, Recife - Pernambuco, Brasil
rita.pedrosa@.estacio.br
[3] *Departamento de Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco Rua Dom Manuel de Medeiros, s/n - Dois Irmãos, 52171-900, Recife - Pernambuco, Brasil*
rita.pedrosa@ufrpe.br
[4] *Departamento de Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco*
Rua Dom Manuel de Medeiros, s/n - Dois Irmãos, 52171-900, Recife - Pernambuco, Brasil
claudio.cristino@ufrpe.br

Abstract. The occurrence of disease outbreaks, especially Dengue, Zika and Chikungunya, is on the increase throughout Brazil, and is currently a significant concern for the Recife-PE region due to the high temperatures. This directly affects public health and the population's quality of life, as well as the city's economy and education. To deal with this challenge, public policies are being implemented, such as awareness campaigns, inspections, and case reports. However, underreporting is a common problem due to the lack of demand for health services and difficulties in accessing medical care, which is reflected in official municipal data. In addition, there are gaps in the treatment of the specificity and cause of the problem and its mitigation. To overcome this inaccuracy, dynamic models such as the SIR model have been widely used in epidemiology. This model, which is based on differential equations, describes the temporal evolution of the susceptible, infected, and recovered classes. In Recife, the ambient temperature shows a strong positive correlation with the infection rates of cases of the diseases, which leads to the intensification of prevention campaigns during the summer. A study carried out in 2022 used data from the National Institute of Meteorology (INMET) to adjust a trigonometric function and analyse the seasonal influence of climate on infection rates, applying the SIR model. In addition, monthly iterations were carried out using the Runge Kutta numerical method in Simulink™ software. To improve the prediction and qualification of endemic disease cases, long-term memory modelling was used with a Recurrent Neural Network (RNN), validated based on available epidemiological data, obtaining an RMSE error metric of around 0.8 for the three diseases assessed.

Keywords: Epidemiological Modeling, Recurrent Neural Network (RNN), Arbovirus in Recife-PE.

1 Introduction

The Recife-PE region in Brazil is facing a growing public health challenge: the proliferation of arboviruses, especially Dengue, Zika and Chikungunya, which has had a considerable impact on the region. The people most affected are those living in situations of social vulnerability, such as poor basic sanitation, combined with high temperatures and humidity favoring the proliferation of the flie *Aedes aegypti*, vector of these diseases due to extreme events caused by climate change, creating a continuous challenge for public health and local development.[1, 2]

To reduce these impacts, various actions are being taken by the Pernambuco state government, such as awareness campaigns, inspections and case reports. However, the region still has a recurring problem of underreporting, which makes it difficult to obtain data that represents the real scenario. This is because many people do not seek health services when they are ill and access to medical care is limited in some regions. This lack of accurate official data in the city of Recife creates gaps in our understanding and in the effectiveness of measures to control the spread of the virus in the region.

The SIR models, developed by Kermack and McKendrick in 1927, have been applied to various epidemic scenarios. These models estimate, over time, the theoretical number of people susceptible to the disease (susceptible), the number of sick people (infected) and the number of people who cannot transmit the disease (recovered or deceased) in a population. In this way, the use of dynamic models such as the SIR (Susceptible-Infected-Recovered) model has proved to be a valuable tool in epidemiology. Based on differential equations, the SIR model describes the temporal evolution of susceptible, infected and recovered populations, offering a systematic view of the spread of diseases. [3]

In the study by Choi, E. (2016), Recurrent Neural Networks (RNNs) are a class of neural networks particularly suited to dealing with sequential data, making them useful in various epidemiological health data applications. These models are able to capture temporal and sequence dependencies, which is crucial when analyzing disease progression, response to treatments and predicting outbreaks. [4]

In order to overcome this inaccuracy and improve prevention and control strategies, this study proposes a temporal analysis of arboviruses in the Recife-PE region. By combining dynamic epidemiological models such as SIR and Recurrent Neural Network (RNN) techniques, to improve the prediction and qualification of arbovirus cases, this study proposes the use of long-term memory modeling, correlated with the temperature database available from the Instituto Nacional de Meteorologia (INMET), and previous studies, as well as data made available by the Secretaria Estadual de Saúde de PE.

We sought to identify patterns and relationships between disease cases and climatic variables such as ambient temperature to adjust trigonometric functions and analyze the seasonal influence of climate on infection rates by applying the SIR model. Monthly iterations using the Runge Kutta numerical method in Simulink™ software provided valuable insights into the dynamics of disease spread. [5, 6].

2 Climate Change in Public Health in Recife: Increase in Arboviral Diseases and . Socioeconomic Challenges

Climate change refers to the long-term alteration in the Earth's average weather patterns, characterized by a gradual increase in global temperature. This change is mainly caused by the excessive emission of greenhouse gases (GHG) into the atmosphere, such as carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O), mainly from the burning of fossil fuels (coal, oil and natural gas), deforestation and intensive agricultural practices. Understanding the relationship between climate, economic growth and health is fundamental to implementing the measures needed to minimize risks in all sectors of the economy, according to the Intergovernmental Panel on Climate Change (IPCC, 2022).

These changes are triggering a series of increasingly frequent and intense extreme weather events such as droughts, rains and heatwaves, which pose a serious threat to life on Earth, especially human health. A study by [7] emphasizes the importance of considering multi-hazard scenarios when assessing the risk of infectious diseases. Higher temperatures, combined with other climatic disasters, could increase the risks of diseases such as dengue fever, prompting the need for massive awareness and prevention measures ahead of a dengue outbreak.

Arboviral diseases are a group of illnesses caused by viruses transmitted by arthropods, mainly flies, such as: Dengue, Zika and Chikungunya,on each one has its own characteristics. [8]. In this sense, these extreme weather events not only exacerbate existing social inequalities, but also increase the incidence of diseases which are transmitted by the *Aedes aegypti*.

[9] investigated the incidence of these diseases, highlighting the environmental and social factors that influence their spread. The study reveals a significant increase in cases, attributed to factors such as the favorable climate and a precarious urban context, where the lack of adequate sanitation favored the proliferation of the mosquito responsible for transmission.

In 2022, the population of Recife-PE was 1,488,920 inhabitants and the population density was 6,803.6 inhabitants per square kilometer, according to the Instituto Brasileiro de Geografia e Estatística [10]. Recife is the capital of the Pernambuco state. The city has a humid tropical climate, with high temperatures throughout the year and a more pronounced rainy season from March to August. Economically, the region is a hub for industry, services and commerce. Sectors such as information technology, tourism, financial services and the shipbuilding industry have grown significantly in the city.

Recife, despite its growth, faces a number of socio-economic challenges. Problems such as social inequality, poor urban infrastructure and environmental issues resulting from intense urbanization are evident. According to [11], a serious public health problem in the city is the lack of adequate sanitation infrastructure, with only 44.01% of sewage collected. In addition, extreme weather events such as floods and droughts exacerbate these challenges for the aforementioned authors, widening social disparities and disproportionately impacting the most vulnerable populations.

According to the Epidemiological Bulletin released by the State Health Department [12], the state of Pernambuco showed a significant increase of 113.4% in the number of probable cases of dengue and 28.3% of chikungunya in February 2024, compared to the same period last year. Probable cases include both confirmed cases and those under investigation. The combination of climate change and precarious urbanization creates a scenario conducive to the increase of arboviral diseases in Recife, which requires strategies in the direction of public policies, focusing on reducing inequalities and promoting environmental resilience and improving access to basic health services.

According to the IPCC report (2023), the most disadvantaged regions, such as Brazil and its states, will suffer the most from climate change. These areas, characterized by low levels of economic and social development, poor infrastructure and limited access to basic services, are particularly vulnerable to the adverse impacts of climate change.

2.1 Dynamic simulation: Recurrent Neural Networks

Dynamic simulation modeling is an important method for evaluating interventions in complex systems, offering an effective approach for estimating the results of unforeseen interactions and prescribing actions based on hypothetical scenarios. This technique involves creating computer models that replicate the functioning of real systems, allowing the observation of how these systems respond to different conditions and changes in parameters over time.[13–15]

Several authors have made significant contributions to the subject of System Dynamics Simulation. For example, [14], the pioneer of System Dynamics, laid the theoretical and practical foundations for its initial application in industrial processes. Over the years, this methodology has spread to various areas. [13] is a reference in system dynamics in business and organizational contexts, while [16] stands out in the modeling of systems in the environmental and energy sectors. System Dynamics is a powerful tool for modelling the construction of a model of reality, where the elements and their cause and effect interactions are visualized by means of diagrams and graphs in complex computer simulations, providing valuable information on how interventions that facilitate the exploration of various scenarios and the experimentation of alternatives to adjust the behaviour of the elements within a system systems over time. [17]

Simulations involving arboviral disease data, such as dengue, zika and chikungunya, together with temperature data, are quite complex due to various factors. The analysis requires processing large volumes of historical data and taking into account various climatic factors, such as temperature and precipitation. To address this complexity, it is crucial to use deep machine learning tools, such as Artificial Neural Networks (ANNs). ANNs, a subcategory of Artificial Intelligence, are currently indispensable for understanding and predicting the dynamics of these diseases. [18]

2.2 SIR Model

The history of epidemiology is linked to history itself, starting with the beginnings of medicine, Hippocrates (458-377 BC), and later, Aristotle (384-322 BC). In 1760, the pioneering work of Daniel Bernoulli (1700-1782) began a new stage of epidemiological analysis. Using innovative mathematical tools, Bernoulli contributed to the dynamics of infectious diseases, the mechanisms of propagation and the impact on public health. [19–21]. In 1906, W. H. Hamer proposed a model that considered the number of susceptible and infected individuals and the rate of contact between them to explain the development of epidemics. Sir Ronald Ross (1908), suggested the existence of a population limit of mosquitoes below which the disease would be eradicated.

The study by Sir Ronald Ross (1908) was fundamental to the creation of the Threshold Theorem, formulated by Kermack and McKendrick (1927), which establishes the relationship between the density of susceptible individuals and the occurrence of epidemics. The aforementioned authors developed the Susceptible-Infected-Recovered (SIR) model, which is one of the most fundamental and widely used models in mathematical epidemiology. Initially, the model worked with algebraic equations, and later improved the model with differential equations to describe the dynamics of infectious diseases in a population, thus, allowing for a more rigorous analysis of the evolution of the disease over time.[22, 23] The SIR model, Equation 1, divides the population into three distinct classes:

- Susceptible (S): individuals who can contract the disease in time (t);
- Infected (I): individuals who can transmit the disease in time (t);
- Recovered (R): individuals who have recovered from the disease and are not subject to new contamination at time (t);
- β is the rate of infection;
- γ is the recovery rate;

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI;\ \frac{\mathrm{d}I}{\mathrm{d}t} = \ \beta SI - \gamma I;\ \frac{\mathrm{d}R}{\mathrm{d}t} = \ \gamma R \tag{1}$$

The Susceptible (S) rate class is negative, as it represents the loss of individuals who become infected and change the class at the model, which is why it is negative thus decrease the amount of susceptible class. This rate is proportional to the number of susceptible and infected individuals, and to the contact parameter (β), because the more susceptible and infected individuals in the population, the greater the chance of contact and transmission of the disease. The rate of the Infected(I) class, on the other hand, represents the entry of new infected individuals from susceptible individuals, which is why it is positive. This rate is proportional to the number of susceptible and infected individuals, and to the contact parameter (β), as described in the previous equation. The rate of change of infected individuals is also negative, as it represents the recovery of individuals who become recovered. The loss of infected individuals is proportional to the number of infected individuals and the recovery parameter (γ). [19, 24] In the recovered class rate(R) represents the entry of new recovered individuals from the infected, so it is positive. This rate is also proportional to the number of infected individuals and the recovery parameter (γ). It should be noted that in the SIR model, it is initially necessary to define the initial conditions for each class at the initial time (t = 0), for example: S (0): Initial number of Susceptibles in the population; I (0): Initial number of Infected in the population; R (0): Initial number of Recovered in the population.

3 Methods

The SIR model is based on constant infection and mortality rates as a simplifying hypothesis, which is not valid for regions with a significant increase in temperature during the summer. Therefore, the first stage of the methodology was based on adjusting the ambient temperature of the city of Recife based on the meteorological data provided by the National Institute of Meteorology (INMET) for the year 2022. The reading frequency was considered to be monthly with a sinusoidal characteristic arising from the solstices and equinoxes. The base function for the adjustment was $f(t) = e^{f(T)}$, where $f(T)$ refers to the sinusoid corresponding to the monthly temperature change. Consequently, the function $f(T)$, Equation 2:

$$f(T) = 43.5206 sen(0.002t + 2.4748) + 2.2518 sen(0.6778t + 0.4144) \tag{2}$$

After adjusting the $f(t)$ function, the infection and recovery rates were rectified by coupling the effects of temperature, modeled by the $f(t)$ function, resulting in the following dynamic formulation for the infection rate, Equation 3 e recovery rate, Equation 4.

$$\beta = \beta_0 \times e^{43.5206 sen(0.002t+2.4748)+2.2518 sen(0.6778t+0.4144)} \tag{3}$$

$$\gamma = \gamma_0 \times e^{43.5206 sen(0.002t+2.4748)+2.2518 sen(0.6778t+0.4144)} \tag{4}$$

The Table 1 summarizes the variables and parameters implemented in the SIR model.

Table 1. Parameters and Variable SIR Model

Variable	Means	Unit	Parameters	Means	Unit
S	Susceptible	[People]	T	Temperature	[°C]
I	Infected	[People]	β	Infection rate	[-]
R	Recovered	[People]	γ	Recovery rate	[-]
t	Time	[Month]	β_0	Infection rate initial	[-]
f(T)	Interpolated function	[°C]	γ_0	Recovery rate initial	[-]

From the typology of the SIR model, present in the Equation 1, it can be seen that it is a system of ordinary differential equations, which can be solved numerically using the commercial software Simulink™, available together with MATLAB™. Simulink is based on a block diagram for implementing and solving dynamic models,

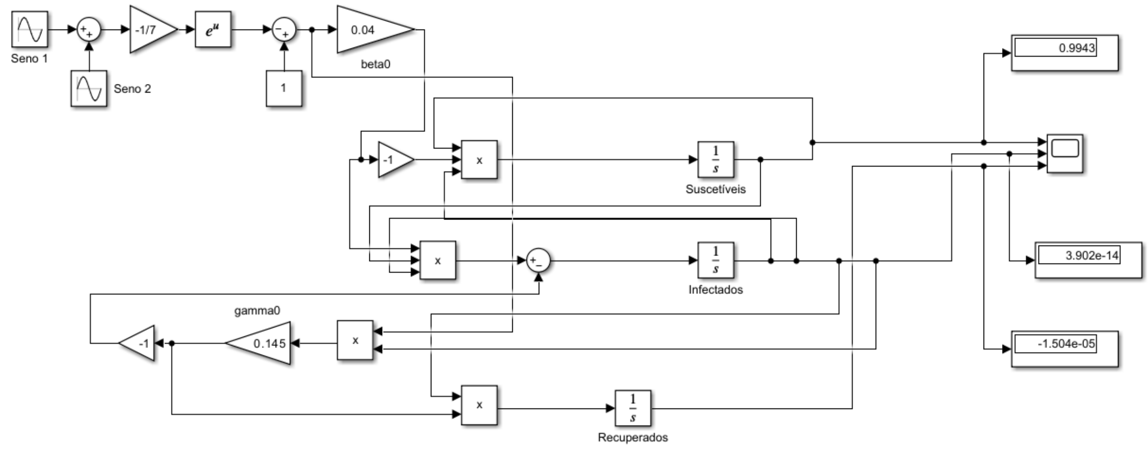

Figure 1. Dynamic simulation diagram

in which each operation can be summarized and interconnected with other blocks. Figure 1 summarizes the information with the application of the SIR model for the city of Recife, the integrator blocks identified as $\frac{1}{s}$ called integrator, solves the differential equations of the system from the fourth order Runge Kutta method, to predict the temporal behavior of the susceptible, infected and recovered.

3.1 Recurrent Neural Network (RNN)

In order to predict the total number of endemic disease cases, the recurrent neural network model was applied using historical data provided by the Recife municipal health department for the year 2022. To predict cases of endemic disease contamination, the Recife health department database for 2022 was used. The database was divided into two proportions: 90% for training and 10% for validation. The algorithms and parameters of the network that were trained are: Adam solver with an initial learning rate of 0.005 and a maximum of 500 epochs and 250 hidden layers for the Long Short Term Memory (LSTM) model, training totaled 1200 iterations with the RMSE (root mean square error) as a comparative metric.

4 Results

In order to qualify and quantify the influence of temperature on the number of cases and consequently predict the number of cases, the SIR and RNN model was applied.

4.1 Model SIR

Figure 2 shows the curves for the three classes of the SIR model, applied on a dimensionless scale as a function of the population of the city of Recife in relation to the iterations. The green curve refers to the susceptible, which has the greatest negative slope just when the ambient temperature is at its maximum, where the reduction in susceptible people is directly observed with the rate of increase from 0.2 to 0.65, at which point the infections are effective. The orange curve below those infected increases to a small extent as a result of the slight increase in the recovery rate where, because a greater number are becoming infected, a proportion will be recovered.

Figure 3 is subdivided in two between the comparison between the curve of observed and predicted cases and the relative prediction errors. The prediction made for 25 months was analyzed in two intervals:

- **Interval I (0 to 13) month**: the prediction resulted in an underestimation of the cases due to the observed curve being higher than the predicted one, an interval which also presented the greatest error. The difference between the values is mainly due to the quality of the data as well as the number of hidden layers needed for a more accurate description of the interval;
- **Interval II (14 to 25) months**: in the second interval, the predictions were more accurate with much smaller relative errors and closer to the curves. One of the observations to be made is that the temperature-corrected infection and recovery rates significantly improved the results obtained in this interval.

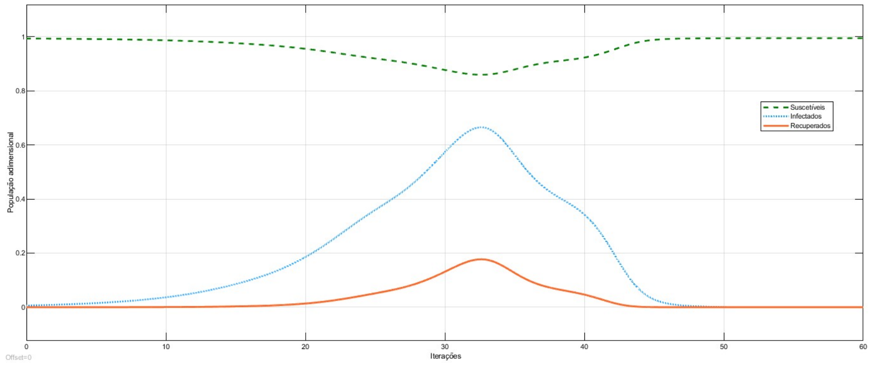

Figure 2. Model SIR Results

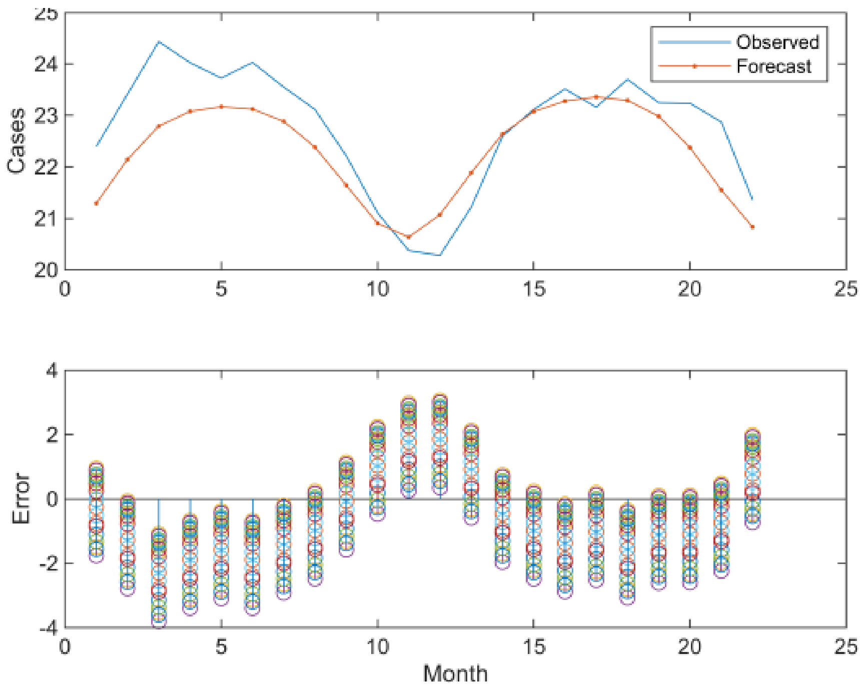

Figure 3. RNN Predictions

5 Conclusions

Endemic diseases such as Dengue, Zika and Chikungunya are an annual health issue and public policies for treatment, prevention, awareness-raising and campaigns aimed at mitigating the social impacts of the diseases. The application of mathematical models to predict the number of cases is highly useful for informing decisions and dimensioning prevention campaigns. This study applied two numerical methodologies: SIR and recurrent neural network, with the difference of incorporating temperature variation as a function of time into the recovery and infection rates, which were originally considered constant. It can be concluded that the models described qualitatively (SIR) and quantitatively (RNN) with an RMSE of around 0.8. It was also found that the predictions with the neural network were more accurate from the second interval examined.

As future improvements, we suggest training the neural network with a greater number of hidden layers, smoothing the original data when pre-processing the data, applying it over a longer time interval and assessing the impact of campaigns on the number of cases.

Acknowledgements. Professora bolsista de produtividade do Centro Universitário Estácio em Recife.

References

[1] de P. Recife. Dados recife. *Secretaria de saúde*, vol. 1, n. 1, pp. –, 2022.

[2] S. Hussain, W. Xuetong, and T. Hussain. Impact of skilled and unskilled labor on project performance using structural equation modeling approach. *Sage Open*, vol. 10, n. 1, pp. 2158244020914590, 2020.

[3] J. Wu, K. Leung, and G. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: A modeling study. *The Lancet*, vol. 75, pp. 399–400, 2020.

[4] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, vol. 9, n. 8, pp. 1735–1780, 1997.

[5] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, vol. 26, n. 6, pp. 855–860, 2020.

[6] I. MathWorks. *Simulink®: Simulation and Model-based Design : Using Simulink®*. Simulink: Simulation and model-based design. MathWorks, 2005.

[7] M. M. Rahman, M. Bodrud-Doza, M. Griffiths, and M. Mamun. Biomedical waste amid covid-19: perspectives from bangladesh. *The Lancet Global Health*, vol. 1, 2020.

[8] J. M. D. C. Ryan Thompson and D. Constenla. A review of the economic evidence of aedes-borne arboviruses and aedes-borne arboviral disease prevention and control strategies. *Expert Review of Vaccines*, vol. 19, n. 2, pp. 143–162. PMID: 32077343, 2020.

[9] E. d. A. S. Rodrigues, I. M. Costa, and S. d. C. Lima. Epidemiologia da dengue, zika e chikungunyia, entre 2014 a 2016, em uberlÂndia (mg). *Hygeia - Revista Brasileira de Geografia Médica e da Saúde*, vol. 14, n. 30, pp. 62–81, 2019.

[10] de Geografia e I. B. Estatística. *Censo Demográfico 2022*. Instituto Brasileiro de Geografia e Estatística (IBGE), 2023.

[11] A. Paiva, A. Rodrigues, A. Amorim, L. Oliveira, and S. Santos. Aplicação do Índice de vulnerabilidade social como ferramenta de apoio à tomada de decisões para o município do recife-pe. *Journal of Hyperspectral Remote Sensing*, vol. 13, pp. 333–347, 2024.

[12] Secretaria de Saúde. Boletim epidemiológico. *Saúde Pública*, vol. 55, n. 11, 2024.

[13] J. Sterman. Business dynamics, system thinking and modeling for a complex world. *http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=013598/(100)*, vol. 19, 2000.

[14] J. Forrester. *Industrial Dynamics*. Students' edition. M.I.T. Press, 1961.

[15] J. Morecroft. *Strategic Modelling and Business Dynamics: A Feedback Systems Approach*. Serious skills. John Wiley & Sons, 2007.

[16] A. Ford. *Modeling the Environment, Second Edition*. Island Press, 2009.

[17] J. P. Walters, D. W. Archer, G. F. Sassenrath, J. R. Hendrickson, J. D. Hanson, J. M. Halloran, P. Vadas, and V. J. Alarcon. Exploring agricultural production systems and their fundamental components with system dynamics modelling. *Ecological Modelling*, vol. 333, pp. 51–65, 2016.

[18] S. Haykin. *Redes Neurais: Princípios e Prática*. Bookman Editora, 2001.

[19] de P. R. Almeida. Sir epidemic models, continuous and discrete, and vaccination strategies. Dissertação (mestrado em Álgebra; análise; geometria e topologia; matemática aplicada), Universidade Federal de Viçosa, Viçosa, 2014.

[20] M. Sabeti and C. A. R. Castilho. *Modelo Epidêmico Discreto SIR com estrutura etária e aplicação de vacinação em pulsos e constante*. Tese (doutorado), Universidade Federal de Pernambuco, Recife, 2011.

[21] P. White. Mathematical models in infectious disease epidemiology. *Infectious Diseases*, vol. 1, pp. 49–53.e1. Epub 2016 Aug 12, 2017.

[22] H. Yang. *Epidemiologia matemática: estudos dos efeitos da vacinação em doenças de transmissão direta*. Coleção "Livro-texto.". Editora da Unicamp, 2001.

[23] R. Isea and K. E. Lonngren. On the Mathematical Interpretation of Epidemics by Kermack and McKendrick. *General Mathematics Notes*, vol. 19, n. 2, pp. 83–87, 2013.

[24] T. Harko, F. S. Lobo, and M. Mak. Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. *Applied Mathematics and Computation*, vol. 236, pp. 184–194, 2014.