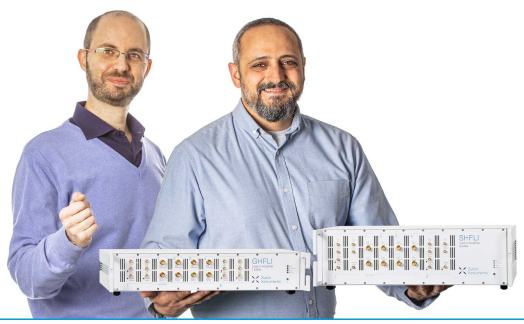
Dynamical systems induced by reaction networks with application to epidemiological outbreaks

Cite as: AIP Conference Proceedings **2522**, 020001 (2022); https://doi.org/10.1063/5.0100921
Published Online: 26 September 2022

M. Lazarova, S. Markov and A. Vassilev



Dynamical Systems Induced by Reaction Networks with Application to Epidemiological Outbreaks

M. Lazarova,[1, a)] S. Markov,[2, b)] and A. Vassilev[3, c)]

[1)] *Faculty of Applied Mathematics and Informatics, Technical University of Sofia, 8 Kl. Ohridski Blvd., Sofia, Bulgaria*
[2)] *Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria*
[3)] *Dill Analytics, St. Kliment Ohridski University of Sofia, Bulgaria*

[a)] *Electronic mail: meglena.laz@tu-sofia.bg*
[b)] *Electronic mail: smarkov@bio.bas.bg*
[c)] *Corresponding author: avassilev@fmi.uni-sofia.bg*

Abstract. We discuss several familiar dynamical systems induced by reaction networks used for the modeling and simulation of epidemiological outbreaks. We are especially interested in dynamical systems that are generated by reaction networks including specific basic reactions such as exponential radioactive decay, logistic or Gompertz growth, *etc.* We explain how reaction networks are "translated" into systems of ODEs. We then formulate certain mathematical properties of the solutions to these dynamical systems and visualize these solutions. We finally present a computational framework for estimating systems of ODEs induced by reaction networks. The main purpose of our work is to demonstrate the practical use of reaction networks in construction of mathematical models describing the time evolution of epidemiological processes.

INTRODUCTION

This work applies the tools of reaction network theory (RNT) to the study of epidemiological processes. We demonstrate how various epidemiological models can be obtained through defining an appropriate reaction network. The studied dynamical systems rely on elements of the Gompertzian and logistic type growth models [1, 2] as their constituent components.

The RNT approach suggests that the classical susceptible-infected-removed (SIR) model, based on the logistic reaction mechanism, describes well epidemic events related to diseases spreading via a "one-to-one" contact pattern between individuals. On the other hand, the two-step exponential growth-decay (2SED) model can be used to simulate epidemic data coming from non-communicable diseases [3]. Our comparative analysis naturally suggests the formulation of a SIR-type model which is situated between the classic SIR model and the 2SED model.

In this new model, referred to as G-SIR, the logistic "one-to-one" contact mechanism is replaced by a catalytic (Gompertzian) one. The proposed G-SIR model can be considered as an intermediate step between the SIR and the 2SED models. We illustrate graphically the shapes of the solutions to the three discussed models and formulate a hypothesis that relates the different model reaction mechanism to the contact patterns of the particular disease. We also present a computational framework for estimating systems of ODEs induced by reaction networks and apply it in a Monte Carlo simulation to estimate a particular version of the G-SIR model.

The paper is organized in the following way. The next section introduces briefly the main concepts of RNT and presents several popular RNT models that can be used as building blocks in epidemiological models. We then formulate the epidemiological models themselves, including the G-SIR model, present their mathematical characterization and visually illustrate the shape of their solutions. The penultimate section turns to estimation and develops the computational framework needed to estimate reaction network models, along with an application to the G-SIR model. The final section concludes.

PRELIMINARIES AND CLASSICAL RNT MODELS

Systems of reactions—chemical, biological or social—can be formalized as *reaction networks* and the study of reaction networks is referred to as *reaction network theory* or, in keeping with established tradition, *chemical reaction*

Application of Mathematics in Technical and Natural Sciences
AIP Conf. Proc. 2522, 020001-1–020001-11; https://doi.org/10.1063/5.0100921
Published by AIP Publishing. 978-0-7354-4361-7/$30.00

network theory. Reaction networks are symbolically presented as systems of elementary reactions and usually are presented in the form

$$S+Q \xrightarrow{k} P+R.$$

The interpretation of this expression is that two or more species on the left side of the arrow, called *reactants* or *reagents* (in this example species S and Q) react and as a result of the reaction one, two or more species, named *products* (here P and R) are produced. The considered reaction network involve one or more decaying species interpreted as environmental resource [3, 4].

All species (reactants and products) partaking in a reaction are denoted by uppercase letters. A positive number called "rate parameter" is written over the reaction arrow. It indicates the velocity of the reaction. The reactants on the left side of the reaction arrow either decay or remain constant, whereas the product species on the right side of the arrow are growing. In some cases species may appear two or more times at one side of the arrow, such as $A+A$, briefly denoted as $2A$. The sign "+" has different meaning when it is placed on the left or on the right side of the reaction arrow. The concentrations (masses) of the species are taken as functions of time t and are denoted by corresponding lowercase letters, such as $s = s(t), p = p(t), \ldots$ The species are assumed to be involved in a reaction network which is governed by mass action kinetics. Therefore the reaction network induces an unique dynamical system of reaction equations for the rates $s' = ds(t)/dt, p' = dp(t)/dt, \ldots$ of the concentrations.

Growth-decay Models and Their Applications

The Gompertz Growth-decay Model

The Gompertz growth-decay model is used in numerous applications. Various formulations of the Gompertz model can be found in the literature. For this model the reaction network for species S, X, [2] has the following expression:

$$S \xrightarrow{\nu} P \quad S+X \xrightarrow{k} 2X+S.$$

The first reaction shows that species S decays exponentially, being consumed by an "outer" species P. The second reaction suggests that species S serves as a catalyst in the growth process $X \xrightarrow{\nu} 2X$.

The induced dynamical system (via mass action kinetics) for the concentrations s, x of the species S, X with the initial conditions $s(0) = s_0 > 0,\ x(0) = x_0 > 0$ is given by $s' = -\nu s,\ x' = ksx$.

The Gompertz model is characterized by a conservation relation. It takes the form:

$$\gamma s + \ln x = \ln c,\ \ln c = \gamma s_0 + \ln x_0,\ \gamma := k/\nu.$$

The Gompertz growth function is often defined as a solution $x = x(t)$ of the following differential equation:

$$x' = \nu x(\ln c - \ln x),$$

where $\nu > 0$ and c are parameters [11]. Figure 1 shows an example of the shape of the solution to the Gompertz model.

For $c = 1$ $(\ln c = 0)$ the solution is given by $s = s_0 e^{-\nu t},\ x = x_0{}^{e^{-\nu t}}$.

The logistic Growth-decay Model

For the logistic growth-decay model the reaction network for species S, X, [2] is given by

$$S+X \xrightarrow{k} 2X,$$

where $k > 0$ is the reaction rate and $2X$ is an abbreviation of $X+X$.

The induced dynamical system (via mass action kinetics) for the concentrations s, x of the species S, X with initial conditions $s(0) = s_0 > 0,\ x_0 = x_0 > 0$ is given by $s' = -ksx,\ x' = +ksx$.

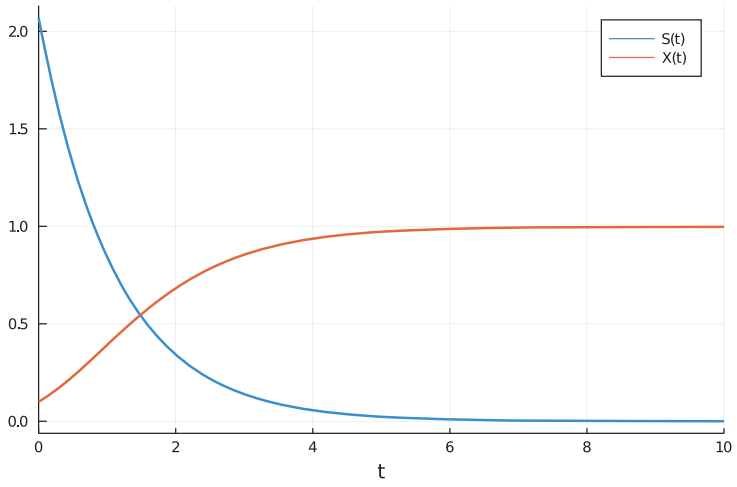


FIGURE 1. The solution to the Gompertz growth-decay model for $k = 1$, $v = 0.9$, $s_0 = 2.07$, $x_0 = 0.1$

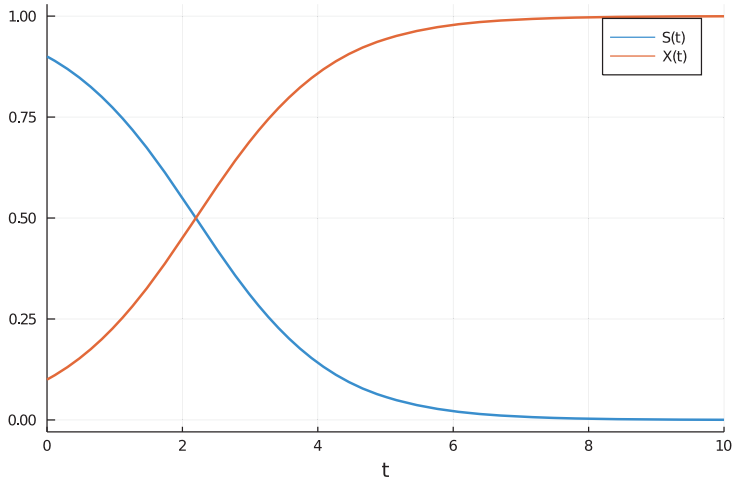


FIGURE 2. The solution to the logistic growth-decay model for $k = 1$, $s_0 = 0.9$, $x_0 = 0.1$

The conservation relation for the model is

$$s' + x' = 0 \quad \Rightarrow \quad s + x = c = \text{const.}$$

In addition, the dynamical system implies that

$$s' = -ks(c-s),\ x' = kx(c-x), c = s_0 + x_0.$$

Assuming $c = 1$ and $s_0 = x_0 = 1/2$, the solution is

$$s(t) = e^{-kt}/(1+e^{-kt}),\ x = 1/(1+e^{-kt}).$$

Figure 2 shows the solution to the logistic growth-decay model.

Compared to the Gompertz model (for appropriate parameterization to ensure comparability), the logistic model exhibits faster growth of the product species (Figure 3). This suggests that in epidemiological applications the logistic model may be used in cases where transmission takes place as a result of direct contact. In contrast, the Gompertz model may be more appropriate in situations where transmission happens indirectly and therefore the spreading of the disease may be slower.

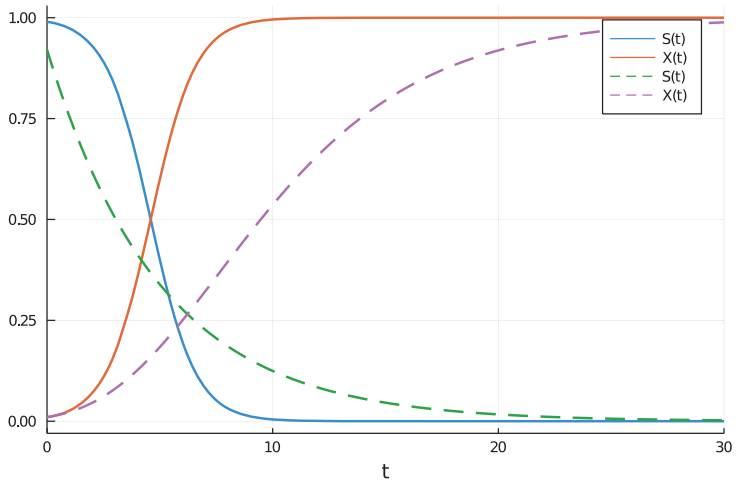


FIGURE 3. Comparison of the solutions of the logistic and Gompertz models. Logistic model: solid lines, $k = 1,\ s_0 = 0.990,\ x_0 = 0.01$. Gompertz model: dashed lines, $k = 1,\ \nu = 0.2,\ s_0 = 0.921, x_0 = 0.01$

One-step Exponential Growth-decay Model

The one-step exponential growth-decay model is defined by the reaction network

$$S \xrightarrow{k} P.$$

The dynamical system generated by this reaction network is $s' = -ks,\ p' = +ks$. This system is coupled with the initial conditions $s(0) = s_0 > 0,\ p(0) = p_0 \geq 0$ to obtain an initial value problem. The one-step exponential growth-decay model is characterized by the conservation relation

$$s' + p' = 0 \quad \Rightarrow \quad s + p = c = \text{const},\ s_0 + p_0 = c.$$

The solution to the model is the following:

$$s(t) = s_0 e^{-kt},\ p(t) = c - s_0 e^{-kt},\ c = s_0 + p_0.$$

It is graphically illustrated in Figure 4.

EPIDEMIOLOGICAL MODELS AND APPLICATIONS

The classical RNT models presented in the previous section provide the required building blocks to move towards modeling epidemiological phenomena. The basic paradigm for transitioning between different compartments or stages of a disease is provided by the two-step exponential growth-decay model, a generalization of the one-step exponential growth-decay model. The 2SED model thus represents the most direct, linear type of transition between stages where interactions do not play a role.

The epidemiological SIR model constitutes a further generalization of the 2SED model with an interaction mechanism added. It embeds a logistic reaction in its formulation and is suitable for situations where a disease spreads through a direct, one-to-one contact pattern.

The differences between the logistic and Gompertz models suggest that we can modify the epidemiological SIR model by replacing the logistic reaction with a Gompertz one. This modification is presented at the end of the present section.

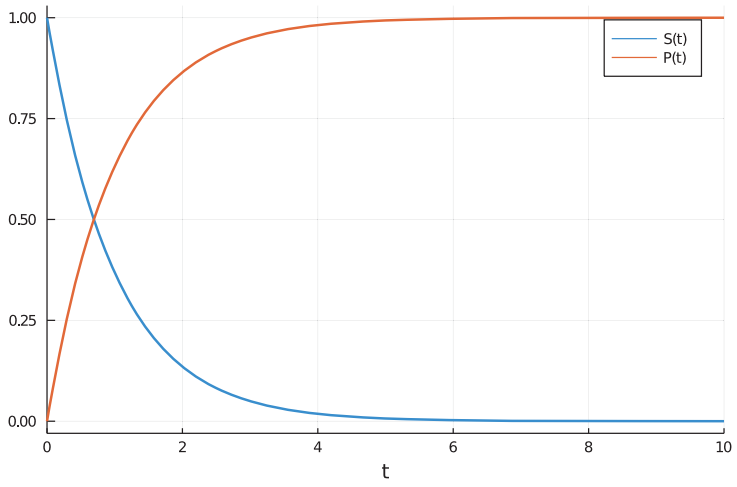


FIGURE 4. The solution to the one-step growth-decay model for $k = 1$, $s_0 = 1.0$, $p_0 = 0.0$

Two-step Exponential Growth-decay Model

The two-step exponential growth-decay model (see [3] for a modern presentation or [12] as a historical reference) is defined by the following reaction network:

$$S \xrightarrow{k_1} P \xrightarrow{k_2} Q.$$

The dynamical system associated with the above network is $s' = -k_1 s,\ p' = k_1 s - k_2 p,\ q' = k_2 p$. The initial conditions required to construct an initial value problem are $s(0) = s_0 > 0,\ p(0) = p_0 = 0,\ q(0) = q_0 = 0$. The conservation relation

$$s' + p' + q' = 0 \quad \Rightarrow \quad s + p + q = c = \text{const},\ s_0 + p_0 + q_0 = c.$$

is valid for the two-step exponential growth-decay model. The solutions are as follows:

$$s(t) = s_0 e^{-k_1 t}$$

$$p(t) = \begin{cases} \frac{s_0 k_1}{k_2 - k_1}(e^{-k_1 t} - e^{-k_2 t}), & k_1 \neq k_2, \\ s_0 k t e^{-kt}, & k_1 = k_2 = k. \end{cases}$$

$$q(t) = \begin{cases} \frac{s_0}{k_2 - k_1}(k_2(1 - e^{-k_1 t}) - k_1(1 - e^{-k_2 t})), & k_1 \neq k_2 \\ s_0(1 - (1 + kt)e^{-kt}), & k_1 = k_2 = k. \end{cases}$$

Figure 5 presents the solution to the two-step growth-decay model for particular parameter values. As visualized on the figure, function $p = p(t)$ can be used for the simulation of an epidemic outbreak [3].

Epidemiological SIR Model

The epidemiological SIR model can be viewed as an appropriate generalization of the two-step growth-decay model. Its reaction network is

$$S + I \xrightarrow{k} 2I, \quad I \xrightarrow{a} R.$$

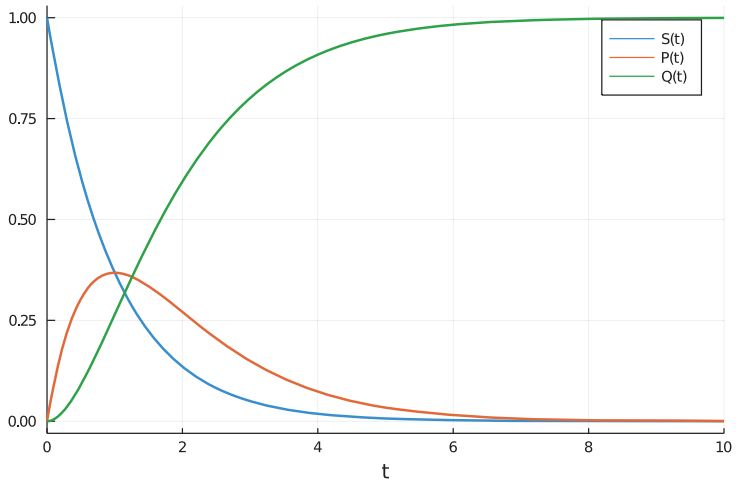


FIGURE 5. The solution to the two-step growth-decay model for $k_1 = k_2 = 1$, $s_0 = 1.0$, $p_0 = 0.0$, $q_0 = 0.0$

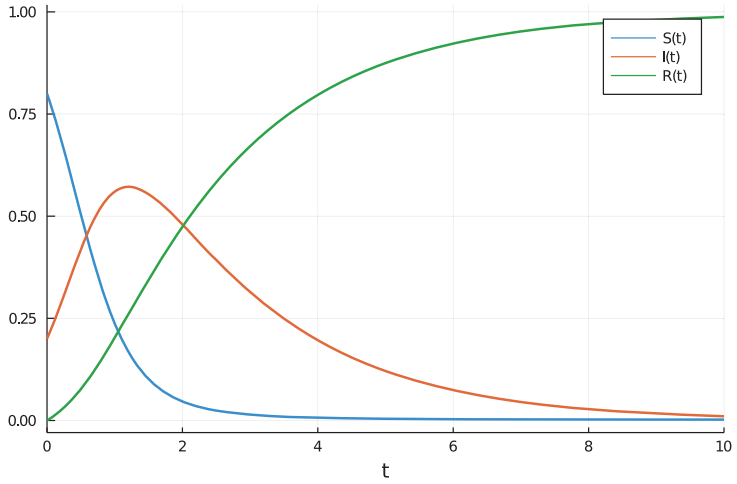


FIGURE 6. The solution to the SIR model in the case of an epidemic outbreak for $k = 3.0,\ a = 0.5,\ s_0 = 0.8,\ i_0 = 0.2,\ r_0 = 0.0$

This reaction network results in the dynamical system $s' = -ksi,\ i' = ksi - ai,\ r' = ai$. The usual initial conditions for the SIR model are $s(0) = s_0 > 0,\ i(0) = i_0 > 0,\ r(0) = r_0 = 0$. The model is characterized by the conservation relation

$$s' + i' + r' = 0 \quad \Rightarrow \quad s + i + r = s_0 + i_0 = c = \text{const.}$$

The interpretation of this reaction network is as follows. There are three classes of species – susceptibles S, infectives I and removed R. The parameters $k, a > 0$ are construed as "infection" and "removal" rate, respectively. An epidemic outbreak occurs when $s_0 > \rho := a/k$. The same condition can be equivalently formulated in terms of the basic reproduction number $R_0^{sir} := s_0/\rho$. Thus, an epidemic outbreak occurs when $R_0^{sir} > 1$. The two situations are illustrated in Figures 6 and 7.

G-SIR Model

The comparison between the properties of the logistic and Gompertz models suggests that we can alternatively embed the respective reaction in epidemiological model formulations. One example of this strategy is a SIR-type epidemiological model where the logistic reaction from the SIR model is replaced by a Gompertzian one. We refer to this new

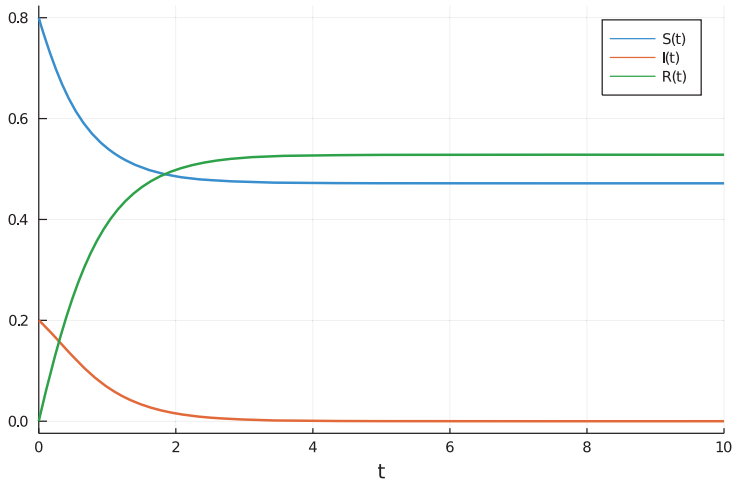


FIGURE 7. The solution to the SIR model in the case of no epidemic outbreak for $k = 3.0,\ a = 3.0,\ s_0 = 0.8,\ i_0 = 0.2,\ r_0 = 0.0$

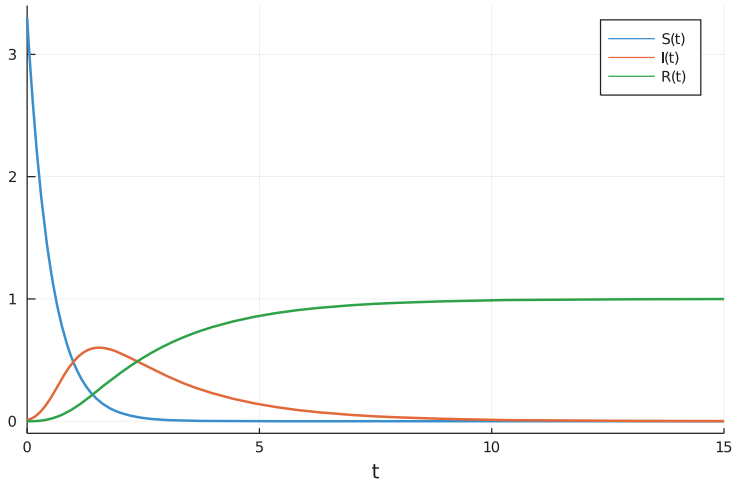


FIGURE 8. The solution to the G-SIR model in the case of an epidemic outbreak for $\nu = 1.93,\ k = 3.0,\ a = 0.5,\ s_0 = 3.3,\ i_0 = 0.01,\ r_0 = 0.0$

model as the G-SIR model [3]. The G-SIR model takes an intermediate position between the SIR and the two-step exponential growth-decay model. The G-SIR model provides an approach to simulate epidemic outbreaks for diseases that are not necessarily communicable according to one-to-one contact spread pattern.

The reaction network for the G-SIR model is

$$S \xrightarrow{\nu} \emptyset, \quad S + I \xrightarrow{k} 2I + S, \quad I \xrightarrow{a} R.$$

This reaction is associated with the dynamical system $s' = -\nu s,\ i' = ksi - ai,\ r' = ai$. Coupling with the initial conditions $s(0) = s_0 > 0,\ i(0) = i_0 > 0,\ r(0) = r_0 = 0$, we obtain an initial value problem for the G-SIR model. Solving the initial value problem yields the following expressions:

$$s = s_0 e^{-\nu t},\ i = i_0 \exp(\gamma s_0 (1 - e^{-\nu t}) - at),\ \gamma := k/\nu.$$

Similar to the SIR model, the G-SIR model can lead to situations where epidemics break out or die out. The first case is illustrated in Figure 8, while Figure 9 shows the case where no outbreak occurs.

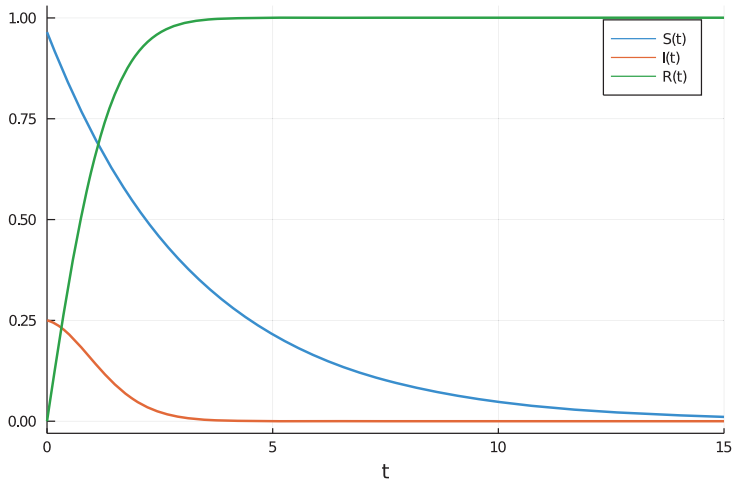


FIGURE 9. The solution to the G-SIR model in the case of no epidemic outbreak for $\nu = 0.3$, $k = 3.0$, $a = 3.0$, $s_0 = 0.965$, $i_0 = 0.25$, $r_0 = 0.0$

RNT MODEL ESTIMATION

The computational results reported in this paper were implemented using the Julia language [5] and various packages from the Julia ecosystem. This allowed us to leverage the language's syntactic expressiveness and speed of execution, while making use of the rich and growing variety of packages available for Julia. All computations were done in the infrastructure provided by the Jupyter Notebook [6].

Here we present an example of how this ecosystem can be employed to automate the process of formulating a reaction network, deriving the associated ODE system, estimating the parameters of the system on data and producing simulations using the estimated model. Our goal is to develop the computational infrastructure for model estimation and testing. To this end, we start by building an estimation pipeline and testing its performance in a Monte Carlo experiment on an artificially generated toy dataset.

Our implementation of the estimation pipeline relies on the Julia packages `Catalyst` [7], `DifferentialEquations` [8], `DiffEqBayes` [9] and `Turing` [10]. These cater respectively for translating a symbolic representation of a reaction network to an ODE system, solving the ODE system, estimation of an ODE system using Bayesian methods and providing the underlying simulation engine for the Bayesian computations.

More specifically, we go through the following steps:

1. Generate test data from an instance of the G-SIR model.
2. Chain together a set of functions that take a symbolic representation of a reaction network, produce the associated ODE system and estimate the rate parameters using Bayesian methods.
3. Apply the above pipeline to the test data and visually check the fit.

To generate the test dataset for the simulation we use an instance of the G-SIR model and perturb it with *iid* Gaussian noise with mean 0 and standard deviation 0.03. The G-SIR model is characterized with the following parameters: $\nu = 1.93$, $k = 3.0$, $a = 0.5$. The initial conditions for the simulation are $s_0 = 3.3$, $i_0 = 0.01$, $r_0 = 0.0$ and the simulation is run over the time interval $t \in [0, 10]$. We sample 50 observations at equally spaced steps over this time interval.

In order to estimate the model in a Bayesian framework, we impose uniform priors over the interval $[0, 5]$ on the three rate parameters $(\nu,\ k,\ a)$. These priors are informative – they guarantee the correct sign of the parameter and, additionally, they impose upper bounds on the admissible parameter values. Thus, a choice similar to ours is appropriate in situations where the researcher has adopted a certain theoretical framework and employs estimation procedures that are guaranteed to be consistent with this framework. The upper bounds imposed by the priors can be made sufficiently high, so as to rule out implausible values without biasing the results of the estimation. At the same time, the chosen priors entail an agnostic attitude on what the likely values inside the support of the distribution are.

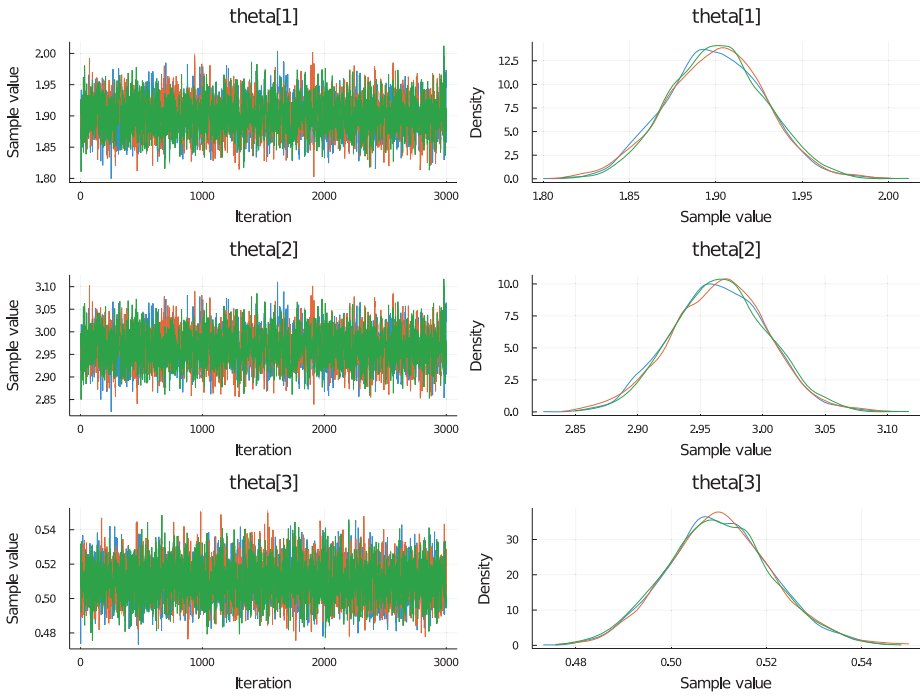


FIGURE 10. Posterior distributions of the main model parameters (theta[1]=ν, theta[2]=k, theta[3]=a)

Parameter	True value	Estimated value
ν	1.93	1.9003
k	3.0	2.9657
a	0.5	0.5101

TABLE I. True and estimated values of the G-SIR model parameters.

To obtain the posterior distributions of the model parameters, we simulate three chains of 3000 observations each from the model. The resulting simulations and distributions are shown in Figure 10. The results indicate adequate convergence and the distributions obtained from the three chains are tightly grouped and similar in shape. Overall, the spread of the posteriors is relatively small, indicating that the available sample is informative enough to obtain estimates of the model parameters.

To get point estimates from the posterior distributions, we use the means of the respective distributions. Table I reports the estimated values and compares them to the true parameter values. The estimates are close to the true values, reflecting an adequate combination of prior specification and sample size to obtain a good approximation to the underlying data generating process.

Finally, we use the estimated model parameters to solve the respective ODE system and check how well the solution fits the data. The results are illustrated in Figure 11. They visually confirm that the estimated model fits the data well. It should be noted, however, that the test case presented here is a relatively simple one in the sense that the data generating process is a well-behaved one with a small amount of noise and the initial conditions for the problem are known. More testing is required to bridge the gap to modeling real-world data.

CONCLUSION

Our work tries to substantiate the claim that the RNT approach provides an intuitive way to specify epidemiological models. Thus, classical epidemiological models can be obtained through the application of RNT tools. Moreover, the RNT approach provide a natural route to modifying the standard epidemiological models in order to obtain new epidemiological specifications. The different versions of the models can be applied to study both diseases with a one-

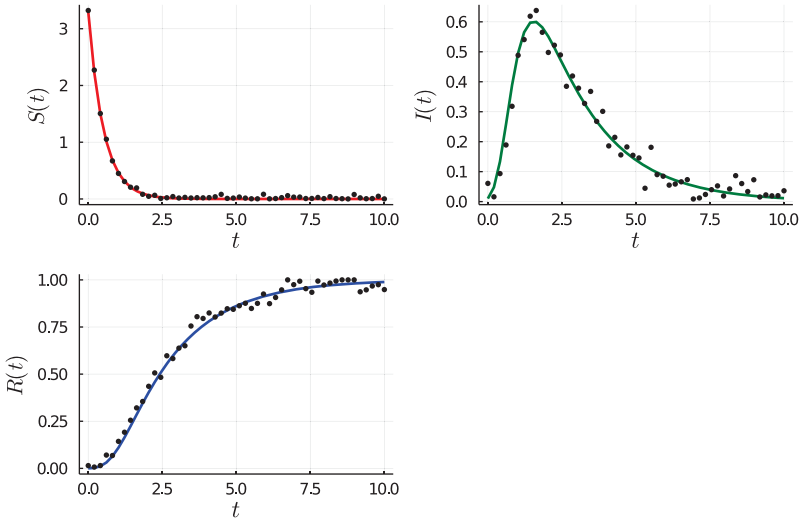


FIGURE 11. Data *vs.* fitted G-SIR curves

to-one contact pattern and non-communicable diseases. The availability of a rich variety of numerical tools facilitates the construction of estimation and simulation pipelines in a flexible and modular manner.

ACKNOWLEDGMENTS

The second author (S.M.) acknowledges support by the National Scientific Program "Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICTinSES)," contract No DO1-205/23.11.2018, financed by the Ministry of Education and Science in Bulgaria.

The authors would like to thank Prof. Pavlina Jordanova for useful questions and comments on the results reported in this paper during its presentation at the AMiTaNS'21 conference in Albena, Bulgaria, 24-29 June 2021 (see the discussion part after references).

REFERENCES

1. R. Anguelov, M. Borisov, A. Iliev, N. Kyurkchiev, and S. Markov, On the chemical meaning of some growth models possessing Gompertzian-type property, Mathematical Methods in the Applied Sciences **41**, 8365–8376 (2018), `https://onlinelibrary.wiley.com/doi/pdf/10.1002/mma.4539`.
2. S. Markov, Reaction network reveal new links between Gompertz and Verhulst growth functions, BIOMATH **8** 1904167 (2019), `http://dx.doi.org/10.11145/j.biomath.2019.04.167`.
3. M. Borisov and S. Markov, The two-step exponential decay reaction network: Analysis of the solution and relation to epidemiological SIR models with logistic and Gompertz type infection contact patterns, Journal of Mathematical Chemistry **41** (2020).
4. G. Lente, "Deterministic kinetics in chemistry and systems biology: The dynamics of complex reaction networks," Springer Briefs in Molecular Science, Springer International Publishing (2015).
5. J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A fresh approach to numerical computing, SIAM Review **59**, 65–98 (2017).
6. T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing, "Jupyter notebooks — a publishing format for reproducible computational workflows, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt (IOS Press, 2016), pp. 87–90.
7. Catalyst.jl (2021), `https://catalyst.sciml.ai/stable/`.
8. C. Rackauckas and Q. Nie, Differentialequations.jl – a performant and feature-rich ecosystem for solving differential equations in Julia, Journal of Open Research Software **5** (2017).
9. DiffEqBayes.jl (2021), `https://diffeqbayes.sciml.ai/stable/`.
10. H. Ge, K. Xu, and Z. Ghahramani, "Turing: a language for flexible probabilistic inference," in *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain* (2018), pp. 1682–1690.
11. Z. Bajzer and S. Vuk-Pavlovic, New dimensions in Gompertzian growth, Journal of Theoretical Medicine **2**, 307–315 (1997).

12. H. Bateman, The solution of a system of differential equations occuring in the theory of radio-active transformations, Proc. Cambridge Phil. Soc. **15**, 423–427 (1910).

DISCUSSION

Question. The Covid-19 pandemic raised the issue of whether the SIR-type models or models based on branching processes are better. Thus, it would be interesting to know if the authors have tried mixtures of cumulative distribution functions as an approach to modeling epidemic phenomena. This approach would permit the modeling of dynamics where multiple local maxima (epidemic waves) arise.

Answer. In the chosen approach the stochastic element is added as a perturbation to the initial differential equation structure. This means that the shapes of the solutions are governed by the properties of the dynamical system and not by the model disturbances. Other approaches are possible but so far the authors have not implemented or tested them.

Question. The work relies on Bayesian analysis in estimating the model parameters. Have the authors tried estimation in a classical framework without imposing specific priors such as the uniform distribution and, if so, which approach leads to shorter confidence intervals?

Answer. Tests were done, again in a Bayesian framework, with alternative prior distributions. Specifically, truncated normal distributions were tested as the alternative to the uniform priors. The posterior distributions obtained were very close to the ones from the simulation using uniform priors. This suggests that the data is informative enough and the likelihood dominates the prior. Thus, the particular choice of prior does not affect the results substantially.

While classical estimation was not done as part of this research, it is certainly a viable option for future work. It should also be noted that the chosen data generating model is relatively well-behaved and does not induce too much noise, which facilitates the discovery of the true parameters.