

On COVID-19 outbreaks predictions: Issues on stability, parameter sensitivity, and precision

M. Stehlík, J. Kiseľák, M. Alejandro Dinamarca, Y. Li & Y. Ying

To cite this article: M. Stehlík, J. Kiseľák, M. Alejandro Dinamarca, Y. Li & Y. Ying (2021) On COVID-19 outbreaks predictions: Issues on stability, parameter sensitivity, and precision, *Stochastic Analysis and Applications*, 39:3, 383-404, DOI: [10.1080/07362994.2020.1802291](https://doi.org/10.1080/07362994.2020.1802291)

To link to this article: <https://doi.org/10.1080/07362994.2020.1802291>



© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC



Published online: 11 Aug 2020.



Submit your article to this journal [↗](#)



Article views: 1325



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



On COVID-19 outbreaks predictions: Issues on stability, parameter sensitivity, and precision

M. Stehlík^{a,b}, J. Kisel'ák^c, M. Alejandro Dinamarca^d, Y. Li^e, and Y. Ying^e

^aLinz Institute of Technology & Department of Applied Statistics, Johannes Kepler University in Linz, Linz, Austria; ^bDepartment of Statistics, University of Valparaíso, Valparaíso, Chile; ^cInstitute of Mathematics, Faculty of Science, P.J. Šafárik University, Košice, Slovakia; ^dCentro de Micro-Bioinnovación, Escuela de Nutrición y Dietética, Facultad de Farmacia, Universidad de Valparaíso, Valparaíso, Chile; ^eDepartment of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa, USA

ABSTRACT

We formulate ill-posedness of inverse problems of estimation and prediction of Coronavirus Disease 2019 (COVID-19) outbreaks from statistical and mathematical perspectives. This is by nature a stochastic problem, since e.g., random perturbation in parameters can cause instability of estimation and prediction. This leaves us with a plenty of possible statistical regularizations, thus generating a plethora of sub-problems. We can mention as examples stability and sensitivity of peak estimation, starting point of exponential growth curve, or estimation of parameters of SIR (Susceptible-Infected-Removed) model. Moreover, each parameter has its specific sensitivity, and naturally, the most sensitive parameter deserves a special attention. E.g., in SIR model, parameter β is more sensitive than parameter γ . In a simple exponential epidemic growth model, parameter b is more sensitive than the parameter a . We also discuss on statistical quality of COVID-19 incidence prediction, where we justify an exponential curve considering the microbial growth in ideal conditions for epidemic. The empirical data from Iowa State, USA, Hubei Province in China, New York State, USA, and Chile justifies an exponential growth curve for initiation of epidemics outbreaks.

ARTICLE HISTORY

Received 26 June 2020
Accepted 21 July 2020

KEYWORDS

COVID-19 outbreaks;
exponential growth model;
SIR model; prediction;
sensitivity

SUBJECT

CLASSIFICATION CODES

Primary 40E10;
Secondary 60G07

1. Introduction

Coronavirus Disease 2019 (COVID-19) produced by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a disease first identified in late 2019 and declared pandemic on March 11. COVID-19 is an international, national and public health emergency [1, 2]. Nevertheless, other important contagious routes such as fecal-oral transmission, has been reported [3]. Fever, cough, sore throat, fatigue, and shortness of breath are characteristics symptoms of COVID-19, nevertheless, other such as, diarrhea, loss of smell and taste, and headache, associated with other organs and systems have been recently accepted. Symptoms may appear 2–14 days after exposure.

CONTACT M. Stehlík ✉ milan.stehlik@jku.at 📧 Linz Institute of Technology & Department of Applied Statistics, J. Kepler University in Linz, Altenberger Strasse 69, Linz 4040, Austria.

© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

According to [4], the experts extracted data regarding 1,099 patients with laboratory confirmed COVID-19 from 552 hospitals in 30 provinces. We can see that 1.18% of the infected people had direct contact with wild animals, 31.30% had been to Wuhan, and 71.80% had contact with people from Wuhan. The median incubation period for the virus is 3 days (range 0–24 days). In addition, studies have found that COVID-19 spreads rapidly from person to person.

Here we formulate selected statistical, mathematical, and real-life challenges of COVID-19 outbreak prediction. In particular we justify an exponential curve from microbiological point of view as a reasonable model for outbreak of COVID-19 epidemics. We need to point out that information criteria for estimation and prediction are not necessarily reaching their maximums/optimums on the same sampling schemes. Such ill posed information relationships can be formulated in the form of 1-st kind Fredholm equations. Under reasonable regularities and simplicity of the underlying process, e.g., autoregressive statistical models, one can apply FIRCEP methodology [5] to obtain such relationships. Such kind of ill-posed information relationships can be formulated also from the perspective of information divergences, and ill-posedness can be translated to normal language as a not-avoidable imprecision of any model with respect to underlying parameter estimation/prediction.

The manuscript is organized as follows. In the following Section 2 we introduce some important information about ill-posed problems. We illustrate ill-posedness on sensitivity of parameter b in a simple exponential growth by using reported data from Iowa State, USA. From economical point of view, for the “restart of country” the correct prediction and estimation of exponential shape of COVID-19 curve plays an important role, this problem has been well visible in Chile. Chile’s economic model is neoliberal doctrine with an important role for the market. The pension system is managed by private operators, the economy is dependent on exports of raw materials such as copper, fishing and agriculture, and medium, small, and micro (family) enterprises. Thus facing the pandemic with strong quarantine measures (e.g., restricted mobility and limited public and private economic activities) is generally a complex problem. According to PAHO (Pan American Health Organization), since the pandemic came to Chile, today (July 1, 2020) there have been monitored 319.493 positive cases and 7.069 deaths. Chile has close to 18 millions of habitants. In this scenario the exponential slope of Chilean COVID-19 growth (positive accumulates cases) can be used as an analytic tool for the proper scaling of governmental policies. On the other hand, the behavior of the exponential growth of COVID-19 in Chile (positive cases) and in particular the slope or rate of contagion, can be used as one of indices of COVID-19 impact on the country’s economy. Since entering the exponential phase, the government has emitted actions and policies of economic aid to face the rise of unemployment to a historic 11.2% (Chilean Institute of Statistics INE) and the fall in the monthly index of economic activity (IMACEC) to $-15,3$ in May 2020 (Banco Central de Chile). IMACEC is predictive statistics of the per capita gross internal product (PIB) in Chile.

Above mentioned societal issues naturally justify the importance to study the stability and sensitivity of underlying dynamics of an individual outbreak models to the input data and estimated parameters. Also, we still have to analyze some special situations; for example, more samples can be detected daily in the later period, which will also cause the growth of number of infected cases. Moreover, the following observations shall be

pointed out: each country is having a different COVID-19 approach and different modeling. Some of countries use discretization of SIR (Susceptible, Infectious, or Removed) model. But not each discretization will be convergent to the same solution of continuous SIR model. Moreover, several effects on equilibrium and stability of SIR has been found, see e.g., [6]. Data from COVID-19 outbreaks are briefly discussed in Section 3. Virological backgrounds for exponential shaped growth curves of COVID-19 outbreak are given in Section 4. In Section 5, we provide a parameter sensitivity study, both from theoretical and empirical perspective, for SIR model without vital dynamics. In the last Section 6, we give concluding remarks and overview of selected important issues for the proper modeling of COVID-19 outbreak.

2. A sensitivity of exponential model to the input/starting parameter

Ill posed problems and parameter estimation are difficult issues for COVID-19 growth models. The random perturbation of parameters can have serious effects on the quality of modeling. As said by Paul Krée in the Preface of [7]: “*Random phenomena has increasing importance in Engineering and Physics, therefore theoretical results are strongly needed. But there is a gap between the probability theory used by mathematicians and practitioners. Two very different languages have been generated in this way...*”

One can indeed observe several discrepancies between COVID-19 policy makers and modelers, possibly caused by the usage of different languages. In principle simple growth models (like the exponential one ae^{bt} or SIR model) looks to be attractive for a straightforward implementation with ad-hoc discretization schemes and various estimation techniques. Thus sensitivity of these models to the principal parameters, e.g. b in case of exponential growth or β in the case of SIR may deceive its user. An independent observer may wonder why the same mistakes in the estimation of outbreaks has been repeated again and again in various countries by using the same models, even when time shift has allowed some possibilities to learn from the mistakes of others. On the other hand we did not want to simplify the whole situation and overemphasize the theory of calibration, estimation and regularization. But more caution is needed in these areas. In the next subsection, we introduce a distributed dynamical system from the stability perspective.

2.1. Distributed dynamical systems

Distributed parameter systems are everywhere. Because they are difficult to deal with, engineers generally avoid partial differential equations. They reason that lumped parameter models will generally suffice and in recent years, finite element analysis has provided a real verification of that idea and the tools to work with. However, there are still some benefits from thinking things in terms of continuum mechanics. The dynamical distributed systems can be very useful for so called inverse problems. Estimation of various flow and mass transport parameters can be seen as the inverse problem of groundwater modeling (see e.g., [8]).

The mathematical term well-posed problem stems from a definition given by Hadamard (see [9]). He believed that mathematical models of physical phenomena should have the properties that

- 1) A solution exists
- 2) The solution is unique
- 3) The solution depends continuously on the data, in some reasonable topology.

Examples of archetypal well-posed problems include the Dirichlet problem for Laplace's equation, and the heat equation with specified initial conditions. These might be regarded as "natural" problems in that there are physical processes that solve these problems. By contrast the backwards heat equation, deducing a previous distribution of temperature from final data is not well-posed in that the solution is highly sensitive to changes in the final data. Problems that are not well-posed in the sense of Hadamard are termed ill-posed. Inverse problems are often ill-posed. Such continuum problems must often be discretized in order to obtain a numerical solution. While in terms of functional analysis such problems are typically continuous, they may suffer from numerical instability when solved with finite precision, or from errors in the data. A measure of well-posedness of a discrete linear problem is the condition number. If a problem is well-posed, then it stands a good chance of solution on a computer using a stable algorithm. If it is not well-posed, it needs to be re-formulated for a numerical treatment. Typically this involves adding some additional assumptions, such as smoothness of the solution. Such a process is known as regularization.

To illustrate mathematical inverse problems, let us consider differentiation. We can construct a simple example with sequence $f_{n,\Delta}(x) = f(x) + \Delta \sin\left(\frac{nx}{\Delta}\right)$, where f and $f_{n,\Delta}$ are the exact and perturbed data. For an arbitrary small data error Δ , the error in the result can be arbitrary large: the derivative does not depend continuously on the data with respect to the uniform norm. Following [10] we have demonstrated some effects that are typical for ill-posed problems, i.e.,

- 1) amplification of high frequency errors;
- 2) restoration of stability by using a-priori information;
- 3) two error terms of different nature, one for the approximation; error, the other one for the propagation of the data error, adding up to a total error;
- 4) loss of information even under optimal circumstances;
- 5) the appearance of an optimal discretization parameter, whose choice depends on a-priori information.

2.2. Parameter stability and sensitivity

Modeling of evolution of infectious disease is important since it can help to predict the future course of an outbreak and to evaluate strategies to control an epidemic. Naturally, models are only as good as the assumptions on which they are based. We believe that parametric modeling is the most common form of modeling used (represented often by a distributed dynamical system). While it is rather straightforward to test the appropriateness of parameters, it can be more difficult to test the validity of the general mathematical form of a model. What is worse, even in the positive case, the sensitivity of the solution to parameter changes (initial conditions included) must be

taken into account. In general, it doesn't really matter if it is deterministic or stochastic, continuous or a discrete model. Similar considerations hold essentially to all of them.

Let X be a smooth manifold. The mapping $\phi: X \times \mathbb{R} \rightarrow X$ ($\phi: X \times \mathbb{Z} \rightarrow X$) is called *the continuous (discrete) C^k -dynamical system* on X if

1. $\phi(x, 0) = x$ for all $x \in X$;
2. mapping $\phi_t: X \rightarrow X$, $\phi_t(x) = \phi(x, t)$ is the C^k -diffeomorphisms for all $t \in \mathbb{R}$ (\mathbb{Z});
3. $\phi_{t+s} = \phi_t \circ \phi_s$ for all $t, s \in \mathbb{R}$ (\mathbb{Z}).

ϕ is often called the evolution function of the dynamical system and X a phase (state) space. Most common construction of dynamical systems is given by initial value problems of ordinary differential equations (or difference equations for discrete case). We illustrate it in the next on typical (evolving in time t) parametrized ODE system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}, t), \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\boldsymbol{\theta} \in \mathbb{R}^p$ is the vector of parameters and $\mathbf{f}: \mathbb{R}^{n+p+1} \rightarrow \mathbb{R}^n$ represents the dynamics. Naturally, initial state is represented by the initial condition¹

$$\mathbf{x}(t_0) = \mathbf{x}_0. \quad (2)$$

Here indisputable fact is that this condition may depend on the parameters, i.e., $\mathbf{x}_0 = \mathbf{x}_0(\boldsymbol{\theta})$. The above representation subsumes the case where the initial condition may itself be seen as a parameter. We assume here that all dependencies are smooth enough to do analysis. Notice that the solution of (2) is parametrized evolution function $\mathbf{x}(\mathbf{x}_0, t_0; \boldsymbol{\theta}; t) = \phi_t(\mathbf{x}_0; \boldsymbol{\theta})$, i.e., a parametrized dynamical system. Notice that in some literature a dynamical system is triple (\mathbb{T}, X, ϕ) , \mathbb{T} is a monoid (usually \mathbb{R} or \mathbb{Z}).

2.2.1. Stability

Here we assume that \mathbf{x}_0 does not depend on $\boldsymbol{\theta}$ (however it might depend on other parameter $\boldsymbol{\beta}$). Solution $\mathbf{x}(t; \boldsymbol{\theta}, t_0, \mathbf{x}_0)$ of given model is stable² if for every (small) $\epsilon > 0$, there exists a $\delta > 0$ such that having initial conditions within distance δ i.e., $\|\mathbf{x}_0 - \mathbf{x}_1\| < \delta$ remains within distance ϵ i.e., $\|\mathbf{x}(t; \boldsymbol{\theta}, t_0, \mathbf{x}_0) - \mathbf{x}(t; \boldsymbol{\theta}, t_0, \mathbf{x}_1)\| < \epsilon$ for all $t \geq t_0$. Notice that δ can depend only on ϵ . That means a resistance to change in time (the trajectories do not change too much under small perturbations). The opposite situation means instability. Typical example of instable system is an exponential growth, which is a natural model of COVID-19 outbreak. Thus the main purpose of developing stability theory is to examine dynamic responses of a system to disturbances as the time approaches infinity. But in practical situations this is not the goal. The predictions we are interested in COVID-19 outbreak models, are short-term predictions, e.g., 2-weeks. These short-term predictions motivate the next notion, sensitivity. Here natural question arises, how the instability relates to sensitivity.

2.2.2. Sensitivity

Sensitivity analysis [11] is used to determine how the parameters of a model influence its outputs, i.e., to study of how the uncertainty in the output can depend on different

sources of uncertainty in its inputs. If the observables are highly sensitive to perturbations in certain parameters then these parameters are likely to be identifiable. There is sometimes hard to answer the question how the magnitude of the sensitivities can be interpreted. What is also important to mention is that in linear case a nonzero right hand side (nonhomogeneous system) might influence sensitivity in contrast to stability. Due to smoothness in sensitivity analysis methods we use so called $n \times p$ matrix of sensitivity functions

$$\mathbf{S} = \nabla_{\theta} \mathbf{x}, \quad (3)$$

which can be understood as a local sensitivity measure. I.e., S_{ij} is related to sensitivity of x_i to parameter θ_j . It can be shown by the chain rule that it satisfies the following ODE matrix system

$$\dot{\mathbf{S}} = (\nabla_{\mathbf{x}} \mathbf{f}) \mathbf{S} + \nabla_{\theta} \mathbf{f} \quad (4)$$

with initial condition $\mathbf{S}(t_0) = \nabla_{\theta} \mathbf{x}_0(\theta)$. It is very helpful especially when the explicit form of solution is not known. I.e., we do need to solve dynamical system to study its sensitivity. \mathbf{S} can also be used to study the evolution of the state covariance matrix of the joint vector of the state and the parameters under the assumption of multivariate Normal distribution and a first-order discretization of Equation (4).

It is often better to explain arising differences as a percentage. Here the elasticity can be used. For simplicity now assume $f: \mathbb{R} \rightarrow \mathbb{R}$. The ratio of the relative (percentage) change in the function's output with respect to the relative change in its input is called elasticity and when considering a smooth function f of a variable at point a it is defined as

$$E_f(a) = \frac{a}{f(a)} f'(a) = \frac{d \ln f(a)}{d \ln a} \lim_{x \rightarrow a} \frac{1 - \frac{f(x)}{f(a)}}{1 - \frac{x}{a}} \approx \frac{\% \Delta f(a)}{\% \Delta a}.$$

Clearly, the elasticity³ can also be defined if the input and/or output is consistently negative (away from zero). The elasticity of a function f is a constant α if and only if the function has the form $f(x) = Cx^{\alpha}$, i.e., a power functions. There exist also generalizations to multi-input-multi-output cases in the literature. We also use notation $E_c f$ for the elasticity of function f w.r.t to parameter c .

2.3 A Malthusian growth model (a simple exponential growth model)

This model is the unique solution of (1) with (2) in the form

$$\dot{x} = b x, \quad (5)$$

$$x(t_0) = a, \quad (6)$$

i.e., $n = 1$, $p = 2$, and $\theta = (a, b)^T$. We admit here only $b > 0$ and $a > 0$.

Remark 2.1. Of course in this case one can obtain all forthcoming information directly from explicit form of the solution

$$x(t) = a e^{b(t-t_0)} \quad (7)$$

since it is known.

From the stability point of view it is clear that here we deal with instability. This follows directly from the fact that derivative is positive. But we are more interested in sensitivity. This is because for COVID-19 outbreak estimation and prediction we have to know how the evolution behaves e.g., in two weeks, not in a long-term periods like twelve months. Sensitivity system of Equation (3) has the form

$$\begin{aligned} \dot{S}_1 &= b S_1, \\ \dot{S}_2 &= b S_2 + x, \end{aligned} \tag{8}$$

with $S_1(t_0) = \frac{\partial x}{\partial a}(t_0) = 1$ and $S_2(t_0) = \frac{\partial x}{\partial b}(t_0) = 0$ following from (6).

We do not need to solve system (8) either. Clearly one can find first integral as follows

$$\frac{dx}{x} = \frac{dS_1}{S_1},$$

which is equivalent to $S_1 = k x$. Moreover, after appropriate multiplication and addition of equations we get $\dot{S}_2 x - S_2 \dot{x} = x^2$, which yields

$$\frac{d}{dt} \left(\frac{S_2}{x} \right) = 1.$$

Thus using initial states we have $S_1 = \frac{x}{a}$ and $S_2 = x(t - t_0)$. Now we want to express the change in the output quantity as a percentage of the nominal value of parameters. Often we can compute

$$E_a x = \frac{a}{x} \frac{\partial x}{\partial a} = \frac{a S_1}{x}, \quad E_b x = \frac{b}{x} \frac{\partial x}{\partial b} = \frac{b S_2}{x}$$

even if we do not know explicit form of a solution. Indeed, thanks to result above we have $E_a x = 1$ and $E_b x = b(t - t_0)$. From percentage sensitivity functions we can conclude

- i) When parameter a is changed by 1%, the state change is also permanently 1%.
- ii) When parameter b is changed by $p\%$, the percentage change of state x increases with time linearly. The change in status is linear function of its nominal value, i.e., $\frac{p}{100} b(t - t_0)\%$.

For the better illustration see also Figures 1–4. To perform sensitivity analyses on the population sizes with respect to the uncertain parameters one can use a full factorial design. One can read a lot from graphical representation of sensitivity indices. In the lower subplot of Figure 5, the sensitivity indices (from package multisensi with `design.args = list(b = c(0.08,0.12,0.16),a = c(82,122,162))`) for the main effects and the first-order interactions at time t are given. Their lengths are normalized and differentiated by colors along the vertical bar. One would might to deduce that at first week the population size is sensitive to the main effect of a . However, the upper subplot illustrates how output quantiles (the extreme (tires), inter-quartile (grey) and median (bold line) output values at all time steps) vary along the time steps. Thus, we can avoid over-interpretation of the sensitivity indices since the variability between simulations is low at these times.

Here we comment on the recent number of COVID-19 infected in Iowa State, USA and we also compare such a real data to the models predicting the number of infections in Hubei Province in China and New York State, USA. By the visual check (see Figure 6) we

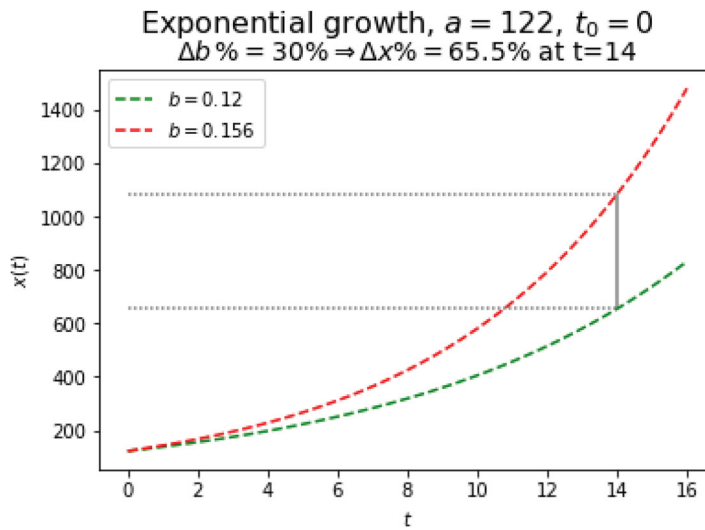


Figure 1. A difference in parameter implies more than a double difference in output.

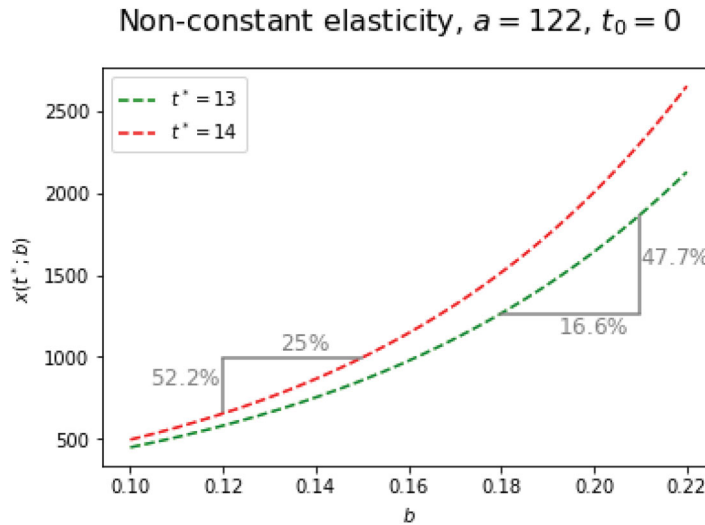


Figure 2. A graphical justification of parameter sensitivity from Figure 1 by the means of elasticity.

adopt an exponential distribution model as a suitable fit for the initial prediction of the number of infected people during the outbreak. Microbiological justification of such model is given in Section 4. The number of people who touched the previous day is used as the starting point and it is calculated according to the exponential model. Several important questions related to challenges of outbreak modeling arisen, namely: What impact does the use of different starting points have? How does the growth rate change in short periods? Are the choices of starting point influential for the growth rate? As follows we will illustrate numerical instability of estimation of starting value b_{start} , naturally, further implementation of exponential nonlinear model $x(t) = ae^{bt}$ will be effected by different b_{start} parameters.

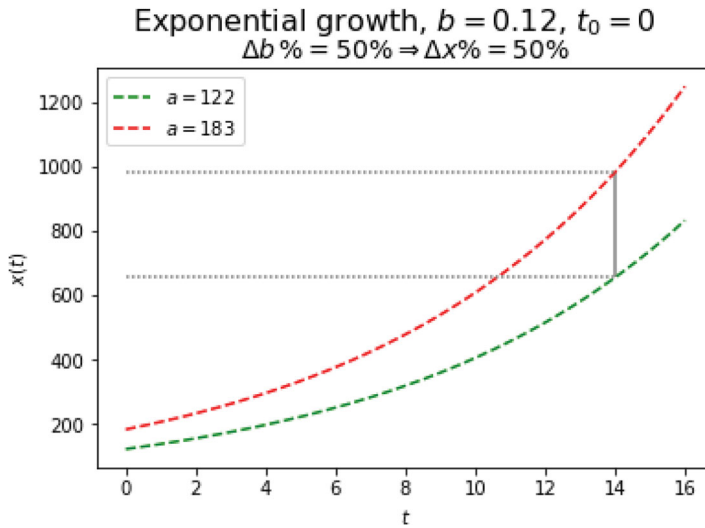


Figure 3. Equality of proportional change of parameter and data output.

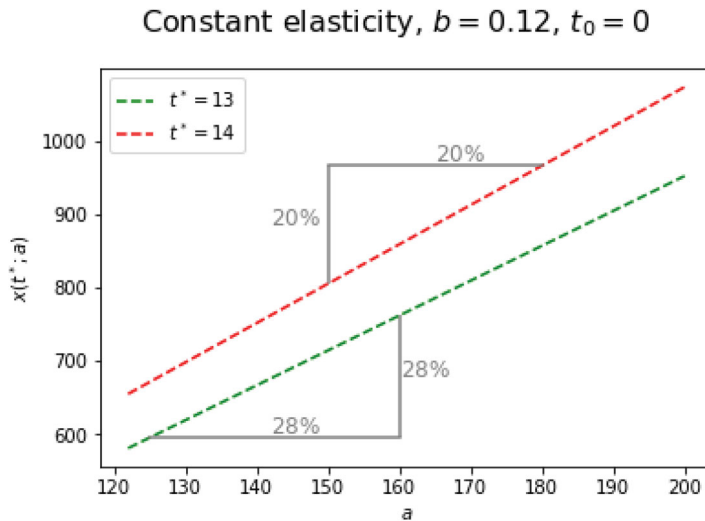


Figure 4. A graphical justification of parameter sensitivity from Figure 3 by the means of constant elasticity.

In the next Figures 7–9, we plotted variability of parameters for the Iowa data. In Figure 7 we plotted the variability of b_{start} in the outbreak period in which cumulatively 300 people were infected. Here we used the previous date as date zero (starting date), thus a_{start} serves as the starting point. At Figure 8a, we plot b_{start} prediction for the nearest 4 weeks, at Figure 8b, we calculated b_{start} for the nearest half-month and at Figure 8c, we calculated b_{start} for the nearest month. The graph on Figure 6 shows that the increase of infected people in Iowa approximately follow the exponential distribution. But if we count different period of time, will the growth rate change a lot? Can we use that to predict a future data? And why the changes happen? On Figure 9 we depict behavior of b_{start} parameter for a longer period from April 25th until May 23rd.

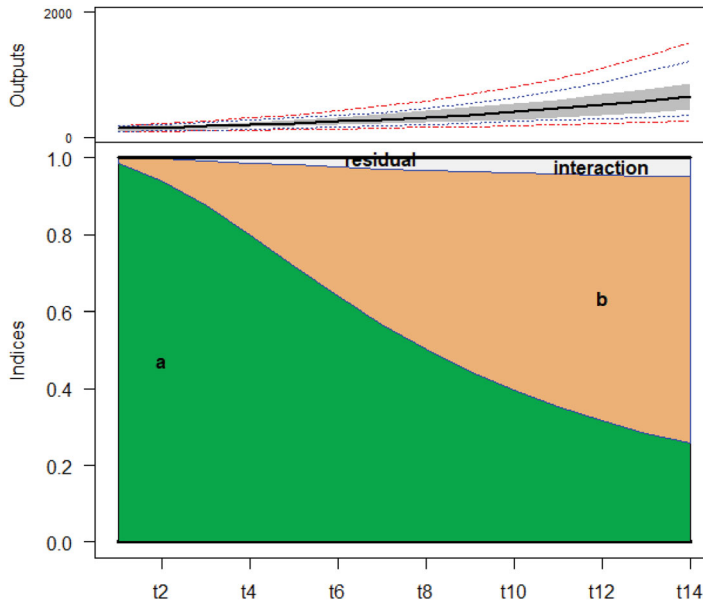


Figure 5. Graphical representation of the sensitivity indices generated by the package multisensi.

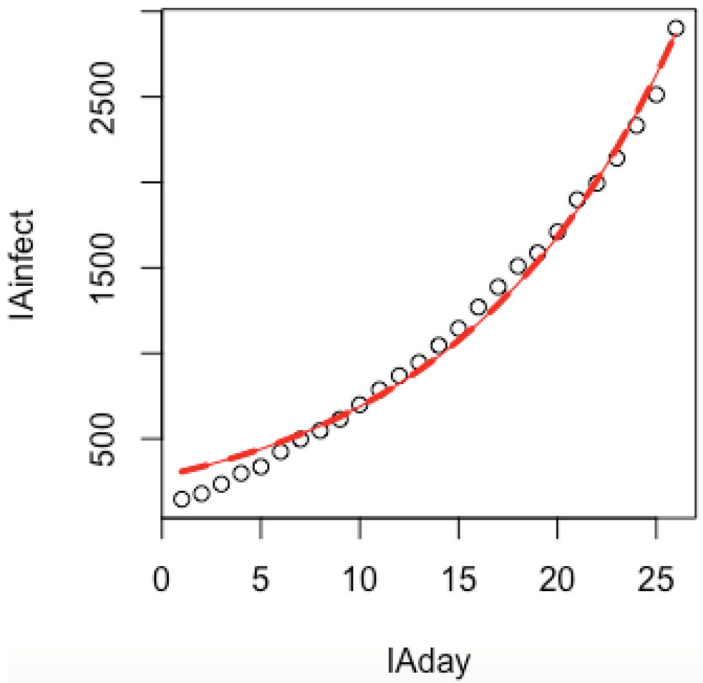


Figure 6. Exponential fitting.

We can conclude from the Figures 7–9 that the growth rate fluctuated a lot even during a relatively short period of time during COVID-19 outbreak in Iowa. But why this happens? We may think about some main reasons: the changes of test approach, increase of tested cases, delays of getting test results, social distance policies, among

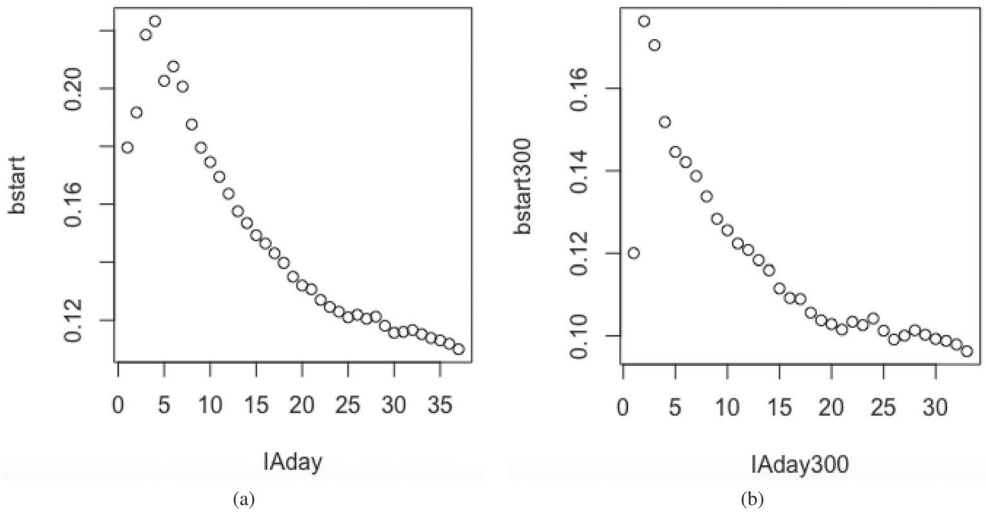


Figure 7. Fluctuation of b_{start} .

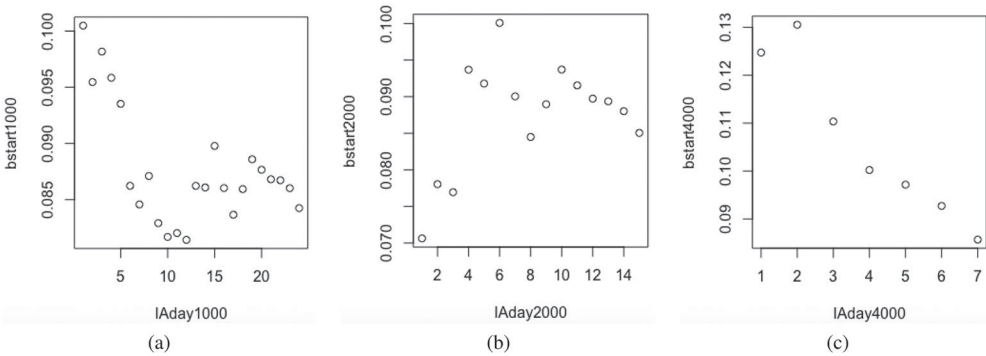


Figure 8. Fluctuation of b_{start} .

others. Moreover, sensitivity of the parameters of the model and its ill-posedness played a fundamental role.

We used the different growth rates from closed 2 weeks to predict for follow-up 2 weeks: Due to that, we can see the choice of growth rate causes the huge differences between individual predictions, the error between two rates is also increasing exponentially, see Figure 10. Summing up, it causes more than 10,000 difference only in 2 weeks, see Figure 11. These graphical exercises illustrate the importance of parameter sensitivity phenomenon in the COVID-19 outbreak models.

3. On COVID-19 outbreaks

The Institute for Health Metrics and Evaluation (IHME) at the University of Washington School of Medicine Study also made predictions of COVID-19 in Iowa. The study conducted before the end of March 2020 said the most recent estimate predicts 349 deaths by August 4 with a peak on May 4. Nationwide the study predicts that

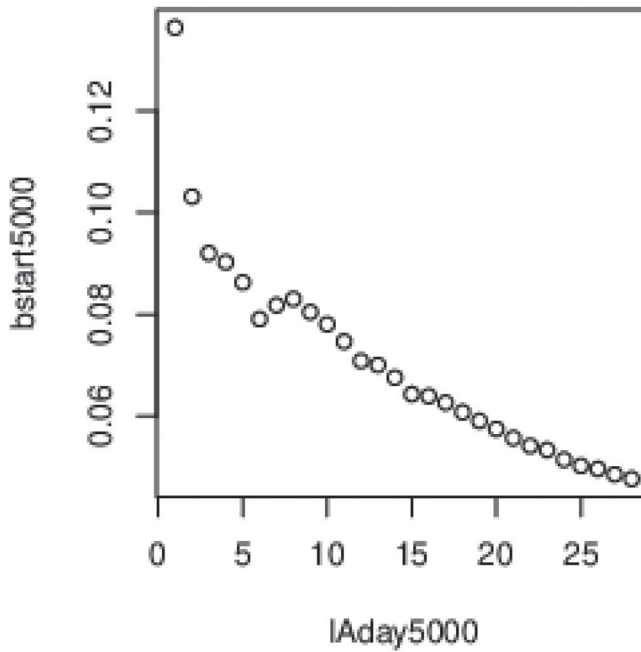


Figure 9. Behavior of b_{start} parameter based on Iowa data from April 25th until May 23rd 2020.

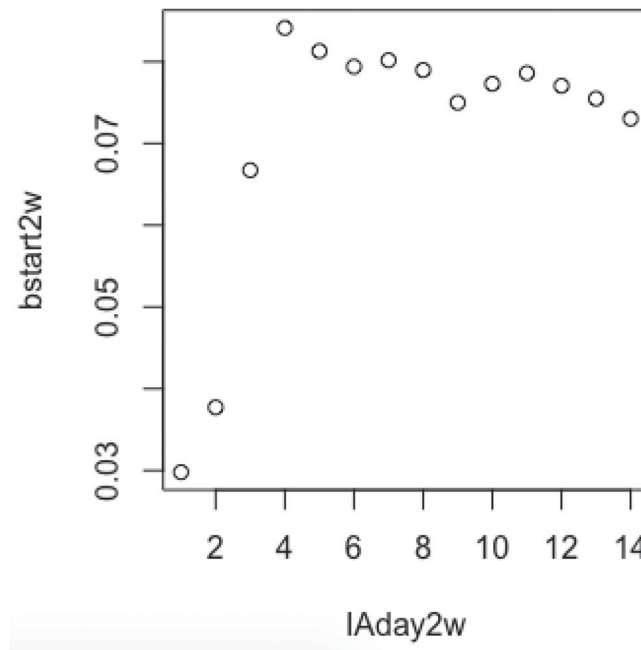


Figure 10. Fluctuation of b_{start} .

74,073 people will die by August 4, which is higher than the study’s previous estimate of 67,641, because longer epidemic peaks in some states and that deaths were not falling as quickly as anticipated.

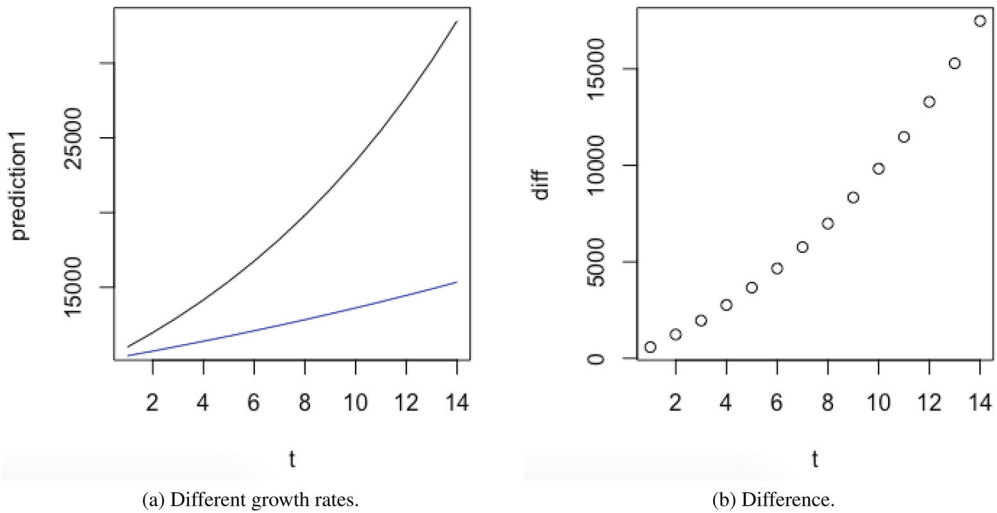


Figure 11. Prediction for future 2 weeks.

Table 1. Age groups with confirmed cases.

	Total	Child	Adult	Middle age	Older adult	Elderly	Unknown
Total	10404	212	4278	3938	1469	505	2
Polk	1875	26	837	578	279	155	0
Black Hawk	1396	21	662	562	122	27	2
Woodbury	1323	66	521	571	158	7	0
Linn	770	19	287	224	163	77	0

We sampled (at the end of March) the populations and the number of confirmed cases and deaths in some counties in Iowa. We found that Polk county was the top one in both population and confirmed cases. But Black Hawk and Woodbury, with populations nearly four times smaller than Polk’s, are in the top three with more than 1,000 confirmed cases. We looked for the median age of several counties (Polk: 35.8; Black Hawk: 34.9; Woodbury: 35.6; Linn: 37.4; Marshall: 38.5; Dallas: 35.1; Johnson: 29.9). So we can conjecture that the number of confirmed cases in each county depends on age. Mortality Rate in Iowa State have been by then 2.1%.

Then we chosen the five counties with the highest number of confirmed cases. We found that adult and middle age groups have the highest rates of infection. In contrast, children and elderly have lower rates of confirmed cases. However, the death rate of elderly was significantly higher than that of other age groups. So, we can conjecture that older people, although they may not be confirmed of large amount like adult group due to the less activities and smaller population, they are also more likely to die from a variety of diseases and weakened immune systems. In fact, older adults and people who have severe underlying chronic medical conditions like heart or lung disease, or diabetes seem to be at higher risk for developing more serious complications from COVID-19 illness. All these factors can be good candidates for covariates of the outbreak models (Tables 1 and 2).

Table 2. Age groups with deaths.

	Total	Adult	Middle age	Older adult	Elderly
Total	219	5	24	87	103
Polk	45	3	1	12	29
Black Hawk	19	0	5	8	6
Woodbury	4	0	0	3	1
Linn	55	0	7	29	19

3.1. Chile: Heterogeneity rules the country

Chile is a geographical and demographical heterogenous country, practically each region is having a different COVID-19 incidence curve, domination is given by Santiago (the most populated, 5.614 million habitants) and Valparaiso. There is a delay in data, namely 3–4 days, because of medical system. Quality of administrative data should be further analyzed. A common aspect is a possible correlation between population densities and incidence values. However, it can be seen that there are gaps in the entry into the increase phase of positive cases. According to the Chilean health authorities the averages of positives for COVID-19 is 8.384 (with a DESVEST of 25,869 (Ministerio de Salud web page, June 7, 2020). The variation can be explained by demographical and geographical aspect in a first instance, nevertheless other factors such as ecology and undiscovered contagious routes can be also considered.

Chilean COVID-19 policies include promoting social distancing and the obligation to use personal protection elements in public areas. Quarantine-bound confinement is implemented gradually and slowly in large cities such as Santiago, considering the incidence and hospital capacity.

With the COVID-19 outbreak we can see that lots of infected cases online are growing almost exponentially. The reason why we use exponential growth to model coronavirus outbreaks is that, based on previous epidemics, the first phase of a pandemic follows exponential growth. Further justifications from the virological perspective are given in the next section 4. The exponential growth function is not necessarily the perfect representation of the COVID-19 growth, however, at the beginning of the covid outbreak it looks to be a reasonable surrogate.

4. Virological backgrounds for exponential shaped growth curves for COVID-19 outbreak

Before we start with microbiological backgrounds, we will illustrate in the following Example 4.1 the exponential growth function.

Example 4.1. Exponential growth function on NY and Hubei data

Graphs on [Figure 12](#) plot the cumulative data of number of infected people and death for Hubei and New York state, and we can see the exponential growth of infected people at the initial stage of outbreak. This motivates an exponential model $x(t) = ae^{bt}$, where x refers to the number of infected people, $a > 0$ is the starting point, $b \geq 0$ is a growth rate and time (days) $t \geq t_0$ (here we set $t_0 = 0$).

In this section we provide microbiological and virological reasons for the exponential growth function. At least three levels of virus spread must be considered in predictive

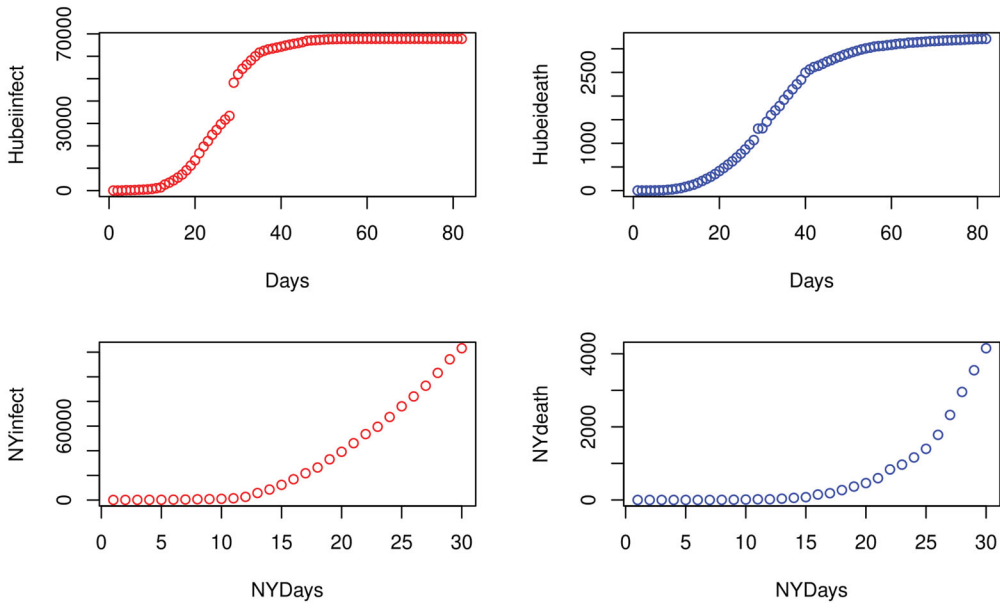


Figure 12. Number of infected and deceased people for Hubei (top) and NY (bottom).

models; cellular level (replicative mechanism), individuals level (organs, anatomy, age, live cycle), and community level (demography, economy, ecosystems).

- **Replication of SARS-COV-2 at the cellular level.** SARS-COV-2 is an enveloped virus containing a single-stranded positive-sense RNA genome. Once the virus binds and enters to the target cell, a complex life cycle mechanism begins using directly the RNAss producing genomic a sub-genomic RNAs. At the same time, a complex mechanism to assemble and releases virions is activated. This process finally destroys the infected cell and spread millions of mature virions.
- **SARS-COV-2 spreading at the individuals level.** Once the virus infects a person, SARS-COV-2 attack an important diversity of cell types affecting different tissues and organs present in the respiratory system, nervous system, digestive system and renal system. This happens because SARS-COV-2 targets tissues whose cells are expressing the widely present Human angiotensin-converting enzyme 2 (ACE2). For this reason, the limited original symptoms were extended to a variety of symptoms ranging from headaches, sore throats, respiratory difficulties, loss of senses (smell and taste), and diarrhea. In this scenario, considering only the original symptoms for medical care or for circulation restriction measures dismisses potential virus disseminators and therefore does not limit the route of transmission. Likewise, since recently has been described that SARS-COV-2 is present in human feces and also has been detected in untreated urban wastewater samples, implies a potential second route of dissemination. The dissemination route currently recognized by the WHO is through the respiratory tract, however, food contaminated by feces (fecal-oral route) and contact with human secretions, among others, should not be ruled out a priori. It is important to consider and understand the rapid and wide spread of COVID-19 pathology in the global population.

4.1. Considerations and assumptions for predictive models of the evolution of COVID-19 in a population

4.1.1. The SARS-COV-2 contagion curves reflect microbial growth under controlled conditions

At the population level the tendencies of dissemination of SARS-COV-2 in each country (measured by the number of daily infected) follow the same trend of a classic microbial growth under controlled conditions. Thus we can consider three phases: a first phase is latency, a second phase is exponential growth, and finally a stabilization phase.

- **Latency phase** It would correspond to the moment when SARS-COV-2 would cross the species barrier and reach man from its natural animal reservoir. From our point of view this event is neither necessarily unique and happening in a single city, nor it could be assumed that SARS-COV-2 has been in the human population since December 2019 to date. The first contact may have been much earlier and not necessarily associated with a single person exposed for the first time to a contaminated animal. The initially proposed scenario can generate inconsistency for any model that is currently used. That is, the so-called patient 0 and the trips to and from China do not necessarily help to explain the spread of the pandemic. Hence, understanding and studying the extent of the virus latency phase is very important, given that it was the time when the virus could have evolved through mutations, thus making humans a new host or reservoir suitable for the replication.
- **Exponential phase.** The contagion trends, of positive SARS CoV 2 to *rtPCR* assay, in each country or city, reflect this phase of microbial growth in optimal conditions. In fact, it is possible to determine a slope, a growth rate and finally establish the maximum rate or rate of infection. For adequate predictive models, in this phase, it is essential to consider asymptomatic infected people or with mild symptoms as vectors of infection. Likewise, it is important to consider that the routes of contagion must be expanded and not necessarily limited to direct contact. For example, faecal-oral contamination should be considered as occurs with foodborne illness. This is very important considering that the virus has been reported to be present in human feces from symptomatic and asymptomatic patients. In fact, the presence of SARS-COV-2 has also been detected in urban wastewater.
- **Stationary phase.** The objective of predictive models is to establish the time of COVID-19 arrival. Beside that proper protective policies can be created based on individuals behavior and governmental actions.

Taking into consideration that the exponential phase indicates the active presence and spread of SARS-COV-2 circulating in a city, region, or country, it is essential to develop models that can help to predict and/or reduce the impact on health systems and local economies. At the same time that they allow verifying the effectiveness of the implemented sanitary measures. A model based on the comparison of growth rates and their duration in a city or country, should take as reference the growth of a virus in its optimal conditions. For this, some analogies and assumptions must be taken into account. In the bacterial

exponential growth model, each bacterium has unlimited nutritional resources, favorable environmental conditions and an adequate metabolism that permit the maximal rate of genome replication and cells division. In spite that viruses do not have metabolisms, like bacteria require replicates its genomes in an adequate environment (human cells and community). In this context, the following assumptions can be made:

- Assumption 1: The current pandemic curves in each country reflect replication of microorganisms in an ecosystem with favorable conditions. When a microorganism such as bacteria find ideal conditions for their development, that is, they are adapted to the environment, they begin an exponential growth based on the duplication of each generation. The doubling time is the time it takes for a generation to double, and this stage of growth is exponential and the slope of the curve at this stage represents the rate of growth. In turn, during the exponential phase, the microorganism will reach what is called the maximum growth rate (μ_{max}) and that is the one in which the doubling occurs in the shortest time during the exponential phase.
- Assumption 2: For SARS-COV-2 to be able to spread in a population, it is required that the individuals provide a metabolic, enzymatic and genetic machinery to produce infectious SARS-COV-2 particles. For this to be effective, individuals must have favorable conditions for viral replication and that involve: age, physiological status, immune status. These conditions should not represent barriers for SARS-COV-2 to replicate. For example, a healthy individual with an active immune system should not be considered as a contagion vector or, “as an adequate substrate for the success in the spread of SARS-COV-2.”
- Assumption 3: The spread implies the interaction between SARS-COV-2 and people, being all the possible routes of contagion and that include, direct contact, fecal-oral contamination by contaminated food, spread in indoor closed spaces (similar to disease outbreaks). In practice, a person can be infected by at least three routes.

Consideration of a Monod-based model (see also [12]). In this sense, and taking into account what Monod proposed in 1947, it is possible to establish the following model proposal.

- Evaluate the kinetics of the exponential phase at the country and city level.
- Develop a statistical methodology to establish whether or not there is a correlation with the population size/density in the country.
- Consider the concept of human vectors to the susceptible population.
- Establish a risk factor to measure, at the city level, the number of people susceptible to becoming ill (symptomatic potentials: diabetics, hypertensive patients, age, pharmacological treatment, immunocompromised, etc.) and the total population of the city.

5. The SIR model without vital dynamics

The SIR model is one of the oldest and simplest of models of an infectious disease in a population that breaks the population into three groups. Birth and death are often

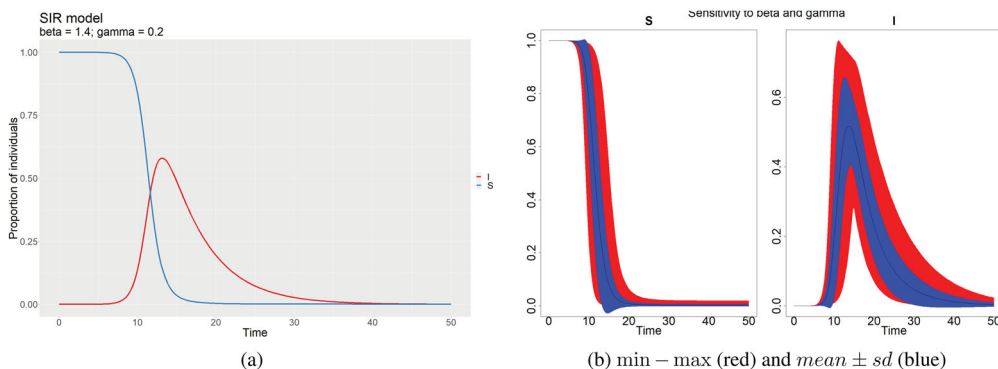


Figure 13. Behavior of S and I and global sensitivity analysis of parameters.

omitted in simple compartmental models since the dynamics of an epidemic, for example, the flu, are often much faster. The SIR system can be expressed by the following set of ordinary differential equations

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS, \\ \frac{dI}{dt} &= \beta IS - \gamma I, \\ \frac{dR}{dt} &= \gamma I,\end{aligned}$$

where S is the proportion of susceptible population, I is the proportion of infected, R is the proportion of removed population (either by death or recovery). Notice that there is not known an explicit form of the solution. However the fact that $S(t) + I(t) + R(t) = 1$ implies that one need only study the equation for two of the three variables.

Remark 5.1. (EXPONENTIAL OUTBREAK). Notice that at the beginning $S \approx 1$ yields $I' = (\beta - \gamma)I = bI$. Thus at the beginning we are basically estimating exponential growth. This can allow us to determine starting condition for parameters discussed in Section 2.3.

We note that the dynamics of the infectious class depends on so-called basic reproduction number $r_0 = \frac{\beta}{\gamma}$. It can be thought of as the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection. It is not a biological constant for a pathogen as it is also affected by other factors such as environmental conditions and the behavior of the infected population. Thus it does not by itself give an estimate of how fast an infection spreads in the population. The most important uses of r_0 are determining if an emerging infectious disease can spread in a population.

We have used packages `deSolve` and `FME` to solve a system of differential equations and to perform a sensitivity analysis. Typical behavior of S and I is plotted on Figure 13(a). The choice of parameters is $\beta = 1.4$ and $\gamma = 0.2$. On Figure 13(b) one can see a global sensitivity analysis of parameters varied over a range $\beta \in [1.2, 1.6]$ and $\gamma \in [0.1, 0.3]$. The effect on model output variables is measured by defining a distribution for each sensitivity parameter and the model is run multiple times. Then envelopes are added to the variables showing the range and $\text{mean} \pm \text{standard deviation}$. In a local

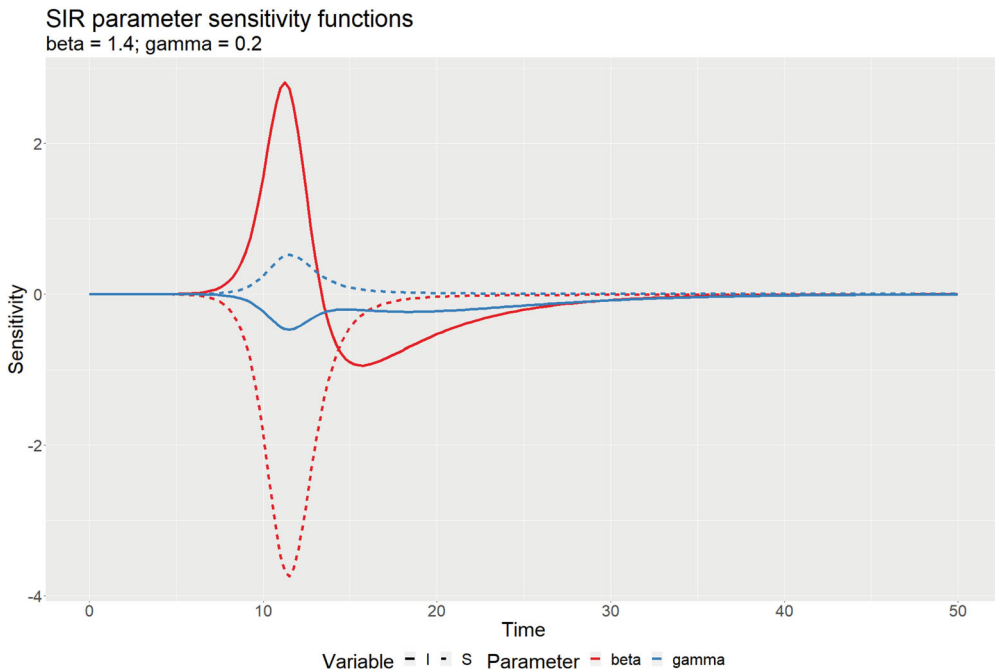


Figure 14. Analysis of local sensitivity of parameters.

sensitivity analysis, the effect of a parameter value in a very small region near its nominal value is estimated. It is good for example, to see which model parameters are more sensitive than the others. From [Figure 14](#) we can deduce that β have the largest values for the sensitivity function, on average, suggesting that this model is most sensitive to this parameter. This is not so surprising. Furthermore, this sensitivity shows peaks on both parameters. This tells us the information about sensitivity of specific times. Finally on [Figure 15](#) we can see how the 1% and 10% increase of β increase the proportion of infected individuals in specific times. Evidently it is not linear.

Now, imagine that we have reliable estimation of parameters β and γ . The question is if the forecast fitting is also qualitatively good. Thus one need a method to compare a goodness-of-fit. We have used an accuracy measure based on percentage (or relative) errors defined as follows:

$$\text{SMAPE}/2 = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{|A_t| + |F_t|} \quad (9)$$

It is the half of symmetric mean absolute percentage error. Actually it is defined as the average of the difference of actual value A_t and the forecast value F_t weighted by their absolute values. The value of (9) is within range of 0% and 100% and thus it is easy to be interpreted. We illustrate this on SIR model fitting for 6 countries. See [Figure 16](#), where on (a) we have used observation of 14 days for fitting (parameters estimation). Clearly in some cases (Iowa, Hubei, Chile) SIR fits quite well since $\text{SMAPE}/2 \in (7\%, 9\%)$. Nevertheless for NY, Slovakia it is worse and $\text{SMAPE}/2 \in (14\%, 29\%)$. However much more important is how the fitting will evolve in the near future. On [Figure 16\(b\)](#) we see the results. Evidently the prediction is not good even for the one week, since there is an increase to

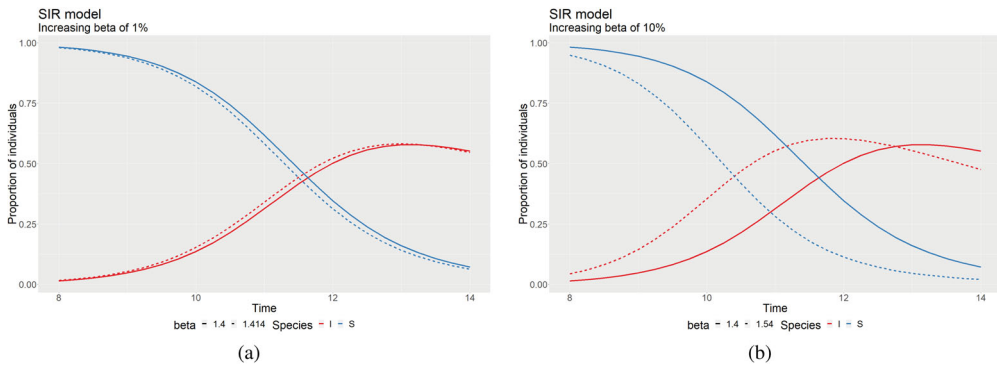


Figure 15. Increase of the proportion of infected individuals.

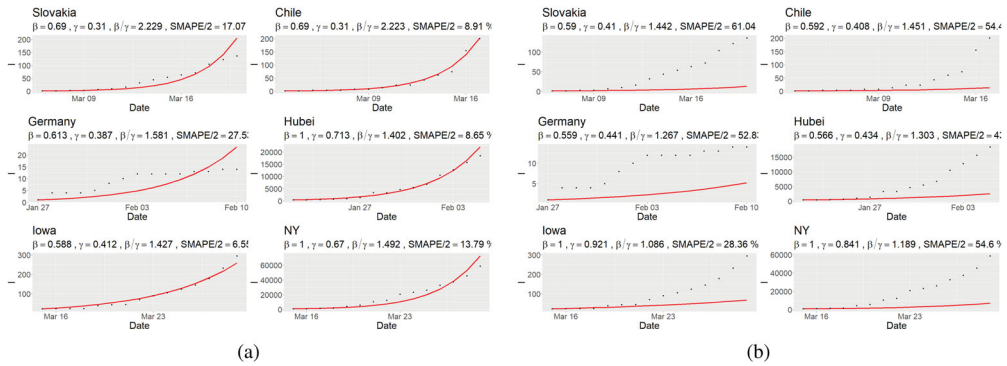


Figure 16. COVID-19 fitted (red) vs. observed (black) incidence.

$SMAPE/2 \in (19\%, 25\%)$ and $SMAPE/2 \in (26\%, 38\%)$, respectively. Thus, SIR model is insensitive even within a short time periods.

6. Conclusions and discussion

The number of infected people is an integer valued mapping $\mathbb{R} \rightarrow \mathbb{N}$. This is typical for modeling a phenomenon evolving in continuous time when the state variable can only take integer values. On the other hand, although we model time as continuum, in practice the measurement of time is always discrete. In particular in the medical sciences, economics etc. time is usually measured, for example, in minutes, days, months or years. One can use positive integers to denote a time. Even if we have short time differences it leads typically to discretization of the continuous model. But what is the effect of discretization (considered as mathematical modeling methodology) on the dynamical behavior of the outbreak model? Can the output change dramatically? What about stability or sensitivity? What about convergence of discretized solution to a continuous one?

One can have a closer look on the switch between discrete and continuous time for the model. From perspective of the fundamentals of dynamical systems see e.g., [13]. Since every continuous dynamical system defines the discrete dynamical system (indeed $\phi: X \times \mathbb{R}$ defines $\tilde{\phi} = \phi|_{X \times \mathbb{Z}}$), a natural question arises, whether something similar holds in the

backward direction. According to the following theorem by Palis (see [13], page 70) we know that there are only a few diffeomorphisms, which can be embedded into a continuous C^0 -dynamical system. In a topological space (X, \mathcal{S}) a set $A \subset X$ is *everywhere-dense* if it is dense in X , i.e., $\bar{A} = X$. Moreover we say that it is *massive* in X , if it contains a set that can be expressed as a countable intersection of open everywhere-dense sets. Notice that every complete metric space is a Baire space, i.e., every its massive subset is an everywhere-dense set.

Theorem 6.1. (J. Palis [13], page 70). *Let X be a smooth manifold. There exist a massive subset G of the set of diffeomorphisms from X onto X such that if $f \in G$ then f cannot be embedded into a C^0 -dynamical system.*

We can deduce from the previous theorem that the class of discrete dynamical systems is “broader” with respect to the structure of their trajectories than the class of continuous dynamical systems.

In [6] authors present four discrete epidemic models with the nonlinear incidence rate using the forward Euler and backward Euler methods. They discuss the effect of two discretizations on the stability of the endemic equilibrium for these models. The sufficient conditions for the stability of the endemic equilibria is established and they emphasize that it can lose stability and the Hopf bifurcation and chaos occur which is not present in the continuous models.

Summarizing, we point out some selected but important aspects for modeling of COVID-19 outbreaks as follows.

- *Quality of the data.* Do we have perfect data for modeling? E.g., number of infected people usually comes from a controlled testing and not from the random sample.
- *Spatiotemporal parameter dependence.* Parameters (e.g., β, γ) might often depend on time or space variables.
- *Real dimension of parametric space.* These parameters can even depend on other factors, conditions or covariates (sex, age, social...), which makes modeling multidimensional. Clearly too few number of parameters can be insufficient, but over-parametrization is also unacceptable.
- *Delayed systems.* Real processes often include aftereffect phenomena in their inner dynamics. Here come time-delay systems on the scene. In the worst case scenarios (time-varying delays, for instance), a lot of caution is needed. Ad-hoc fitting of the models may be potentially disastrous in terms of stability and oscillations.
- *Parameter identifiability.* There is a question whether the parameters of a model can be identified (uniquely or with several solutions) from a specified input-output experiment if perfect data are available. E.g., for linear, time-invariant models, there are several approaches available for this aspect of identifiability analysis. In general identifiability is a property which must be satisfied in order the model can guarantee a precise inference. Usually the model is identifiable only under certain technical restrictions.
- *Discretization* of model (see also the time scale calculus).
- *Non-uniqueness* of the solution of the model and prediction. Bifurcations.
- *Misinterpretation of the results.* E.g., the difference between cumulative and active variable of a dynamical system (active infected individuals vs. cumulative number of infected individuals)

- *Intrinsic and latent heterogeneity.* In a specific country one can define several social groups which can contribute in a heterogeneous way to whole country epidemiological curves.

Notes

1. There can be other types of conditions e.g., several types of boundary conditions, etc.
2. This is stability in the sense of Lyapunov. There exist also notions of other types of stability.
3. Notice, that thanks to logarithm the rules for finding the elasticity of products and quotients are simpler than those for derivatives but sum and subtraction rule does not hold.

Acknowledgments

Authors acknowledge the Slovak Research and Development Agency under the contract no. APVV-17-0568. We acknowledge the professional support of Editor-in-Chief, the unknown Associate Editor and Referees for their constructive comments. The authors are grateful to the bilateral projects Bulgaria – Austria, 2016–2019, ‘Feasible statistical modeling for extremes in ecology and finance’, BNSF, Contract number 01/8, 23/08/2017 and WTZ Project No. BG 09/2017, <https://pavlinakj.wordpress.com/>.

References

- [1] Wang, C., Horby, P. W., Hayden, F. G., Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *Lancet*. 395(10223):470–473. DOI: [10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9).
- [2] WHO. (2020). WHO Director-General’s remarks at the media briefing on 2019-ncov on 11 february 2020. [Internet] World Health Organization.
- [3] Hindson, J. (2020). Covid-19: Faecal-oral transmission? *Nat. Rev. Gastroenterol. Hepatol.* 17(5):259–259. DOI: [10.1038/s41575-020-0295-7](https://doi.org/10.1038/s41575-020-0295-7).
- [4] Guan, W.-j., Ni, Z.-y., Hu, Y., Liang, W.-h., Ou, C.-q., He, J.-x., Liu, L., Shan, H., Lei, C.-l., Hui, D. S., et al. (2020). Clinical characteristics of coronavirus disease 2019 in china. *N. Engl. J. Med.* 382(18):1708–1720. DOI: [10.1056/NEJMoa2002032](https://doi.org/10.1056/NEJMoa2002032).
- [5] Stehlik, M., Kiseľák, J., Bukina, E., Lu, Y., Baran, S. (2020). Fredholm integral relation between compound estimation and prediction (FIRCEP). *Stochastic Anal. Appl.* 38(3): 427–459. DOI: [10.1080/07362994.2019.1696211](https://doi.org/10.1080/07362994.2019.1696211).
- [6] Liu, J., Peng, B., Zhang, T. (2015). Effect of discretization on dynamical behavior of seir and sir models with nonlinear incidence. *Appl. Math. Lett.* 39:60–66. DOI: [10.1016/j.aml.2014.08.012](https://doi.org/10.1016/j.aml.2014.08.012).
- [7] Krée, P., Wedig, W. (1995). *Probabilistic Methods in Applied Physics*. Lecture Notes in Physics Vol. 451. Springer.
- [8] Sun, N. (2013). *Inverse Problems in Groundwater Modeling. Theory and Applications of Transport in Porous Media*. Dordrecht, the Netherlands: Springer Netherlands.
- [9] Hadamard, J. (1902). Sur les problemes aux derivees partielles et leur signification physique. *Princeton Univ. Bull.* 49–52.
- [10] Engl, H., Hanke, M., Neubauer, A. (1996). *Regularization of Inverse Problems. Mathematics and Its Applications*. Dordrecht, the Netherlands: Springer Netherlands.
- [11] Saltelli, A., Chan, K., Scott, E. M., Eds. (2000). *Sensitivity Analysis*. Chichester, UK: Wiley.
- [12] Yin, J., Redovich, J. (2018). Kinetic modeling of virus growth in cells. *Microbiol. Mol. Biol. Rev.* 82(2):1–21. DOI: [10.1128/MMBR.00066-17](https://doi.org/10.1128/MMBR.00066-17).
- [13] Medved’, M. (1992). *Fundamentals of Dynamical Systems and Bifurcation Theory*. Bristol, UK: Adam Hilger.