

# A simple model for COVID-19

Julien Arino <sup>a, b, c, \*</sup>, Stéphanie Portet <sup>a</sup>

<sup>a</sup> Department of Mathematics & Data Science NEXUS, University of Manitoba, Canada

<sup>b</sup> Centre for Disease Modelling, Canada

<sup>c</sup> Canadian COVID-19 Mathematical Modelling Task Force, Canada



## ARTICLE INFO

### Article history:

Received 1 April 2020

Received in revised form 23 April 2020

Accepted 23 April 2020

Available online 28 April 2020

Handling Editor: Jianhong Wu

### Keywords:

COVID-19

Mathematical model

Erlang distribution

Asymptomatic infections

## ABSTRACT

An  $SL_1L_2I_1I_2A_1A_2R$  epidemic model is formulated that describes the spread of an epidemic in a population. The model incorporates an Erlang distribution of times of sojourn in incubating, symptomatically and asymptotically infectious compartments. Basic properties of the model are explored, with focus on properties important in the context of current COVID-19 pandemic.

© 2020 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

As part the authors' work on COVID-19, we have relied several times on a specific model derived from an earlier model (Arino, Brauer, van den Driessche, Watmough, & Wu, 2007). This model, like that previous model, is applicable to a variety of emerging and re-emerging pathogens exhibiting an observable latent period as well as symptomatic and asymptomatic infections. The specificity of the present model is the further incorporation of Erlang distributions of the time of sojourn in some of the important compartments in the model. Indeed, most of the work we have carried out so far on COVID-19 has concerned predictions over a short time period, often no more than a month. In this context, incorporating a better description of sojourn times is extremely important. The present model generalises the simple 3-compartment SIR model to age of infection models, providing a reasonable approximation to the details of progression through infection with a minimal number of parameters and the convenience of an ODE model over integral or PDE models.

In this short note, we present this model and explain some of its features in the context of the current COVID-19 pandemic. We also conduct a simple sensitivity analysis in order to highlight the most important parameters in the model.

The main conclusion of the analysis here is that model responses are highly sensitive to the value of the parameter describing the fraction of cases that are asymptomatic, highlighting the need for intensive research to get a better handle on the value of this critical parameter.

\* Corresponding author. Department of Mathematics, 420 Machray Hall, University of Manitoba, Canada.

E-mail addresses: [Julien.Arino@umanitoba.ca](mailto:Julien.Arino@umanitoba.ca) (J. Arino), [Stephanie.Portet@umanitoba.ca](mailto:Stephanie.Portet@umanitoba.ca) (S. Portet).

Peer review under responsibility of KeAi Communications Co., Ltd.

## 2. The model

We use a simple variation on the classic SLIAR epidemic model for susceptible, latently infected, symptomatic and asymptomatic infectious and removed individuals, with numbers denoted respectively  $S, L, I, A$  and  $R$  (Arino, Brauer, van den Driessche, Watmough, & Wu, 2006). The SLIAR epidemic model has often been used to describe the propagation of diseases caused by virus leading to respiratory illness such as influenza (Jin et al., 2011; Kim, Lee, & Jung, 2017; Li et al., 2020). Contrary to SLIR (or SEIR) models, it allows to consider infection by asymptomatic individuals, which in the case of COVID-19 has been reported to have substantially contributed to disease propagation (Li et al., 2020).

As the time scale of interest is short, the model has no birth or natural death, only death by removal from the infectious compartments  $I$  and  $A$ . It is therefore an *epidemic* model, as opposed to an *endemic* model. Furthermore, because the time horizon for simulations is very short in comparison to reported estimates of incubation period (Backer, Klinkenberg, & Wallinga, 2020; Lauer et al., 2020a) and communicable period (Hu et al., 2020), making a more appropriate description of sojourn times in the incubation, symptomatically infectious and asymptotically infectious compartments is important. Although not ideal, we use an Erlang distribution, i.e., a Gamma distribution with integer shape parameter. To simplify the problem, we use two compartments for each of the  $L, I$  and  $A$  states (Arino, 2020). We could use more if need be; the overall set up would vary very little.

Let us briefly justify this modification in the current context. Consider, for instance, the incubation period, i.e., the time between infection and the onset of symptoms. A wide range of possible durations has been reported. Let us suppose, for instance, that the mean incubation period of COVID-19 is 5.2 days as reported by (Lauer et al., 3AD). Then, comparing the fraction of individuals infected at time 0 and still incubating at time  $t$ , we obtain Fig. 1. The Erlang distribution thus allows both a less pronounced early end and a less extended duration of the incubation period. If need be, we could further extend this behaviour by adding more compartments and thus increasing the shape parameter of the Erlang distribution.

Note that this does not increase the complexity of the model and, for instance, parameter fitting procedures, since we use the same parameter for all compartments in these “chains”; for instance, the incubation period is described using the single parameter  $\varepsilon$ ; the only difference with the exponential case is that, here,  $2/\varepsilon$  is the mean sojourn time in the combined  $L_1$  and  $L_2$  compartments.

The flow diagram of the model is as shown in Fig. 2.

We suppose that incidence takes the form

$$\Phi S = f(\mathbf{x})(I_1 + I_2 + \eta L_2 + \xi(A_1 + A_2))S, \quad (1)$$

where  $\beta$  is the transmission coefficient,  $\eta$  and  $\xi$  are the attenuation factors for transmission by incubating and asymptomatic cases, respectively,  $\mathbf{x} = (S, L_1, L_2, I_1, I_2, A_1, A_2, R)$  is the state vector and  $f: \mathbb{R}_+^8 \rightarrow \mathbb{R}_+$  is the function describing the nature of the overall incidence function. As in Fig. 2, we can also think of incidence as taking the form  $\Phi S$ ;  $\Phi$  in this case is the *force of infection*. Typical choices for  $f$  include  $f(\mathbf{x}) = \beta$ , making the incidence mass action, and  $f(\mathbf{x}) = \beta/\langle \mathbf{1}, \mathbf{x} \rangle$ , with  $\mathbf{1} = (1, \dots, 1)$ , giving proportional incidence.

The system governing the behaviour is then the following:

$$S' = -f(\mathbf{x})(I_1 + I_2 + \xi(A_1 + A_2) + \eta L_2)S \quad (2a)$$

$$L_1' = f(\mathbf{x})(I_1 + I_2 + \xi(A_1 + A_2) + \eta L_2)S - \varepsilon L_1 \quad (2b)$$

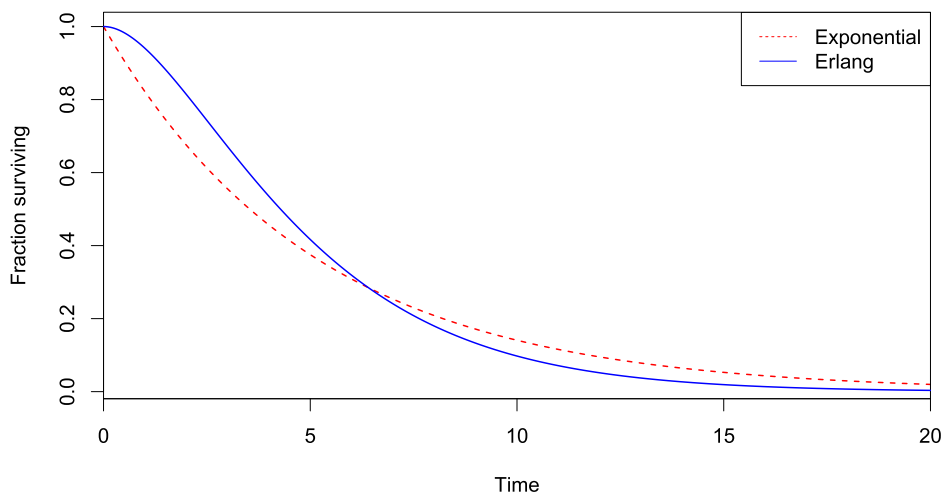
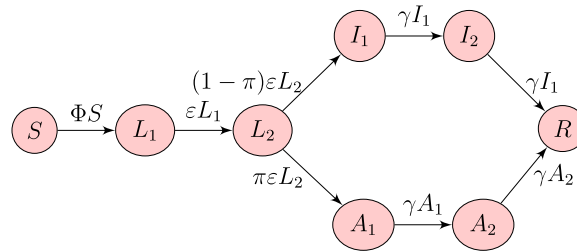


Fig. 1. Comparison of the survival functions of an exponential and an Erlang distributions with mean 5.2 days (and shape 2 in the case of the Erlang distribution).



**Fig. 2.** Flow diagram of the  $SL_1L_2I_1I_2A_1A_2R$  model. Here,  $\Phi = f(\mathbf{x})(I_1 + I_2 + \xi(A_1 + A_2) + \eta L_2)$  is the force of infection.

$$L_2' = \varepsilon(L_1 - L_2) \quad (2c)$$

$$I_1' = (1 - \pi)\varepsilon L_2 - \gamma I_1 \quad (2d)$$

$$I_2' = \gamma(I_1 - I_2) \quad (2e)$$

$$A_1' = \pi\varepsilon L_2 - \gamma A_1 \quad (2f)$$

$$A_2' = \gamma(A_1 - A_2) \quad (2g)$$

$$R' = \gamma I_2 + \gamma A_2. \quad (2h)$$

Note that it is assumed that there can be transmission during the incubation period, as this was reported (Tong et al., 2020). Thus, the compartment  $L_2$  can be interpreted as consisting of pre-symptomatic infectious individuals. A fraction  $\pi$  of individuals is assumed to go into an asymptomatic phase following incubation (and correspondingly,  $1 - \pi$  develop symptoms). Finally,  $\varepsilon$  and  $\gamma$  describe the rates at which incubation and infectiousness end, respectively. By properties of Erlang distributions, the average times spent incubating and infectious (symptomatically or asymptotically) are  $2/\varepsilon$  and  $2/\gamma$  time units, respectively.

### 3. Some properties of the model

The purpose of this paper is not to conduct a thorough mathematical analysis of (2). We refer to (Arino et al., 2006) for considerations on the behaviour of a version of this model with only one of each disease status compartments. It is useful, though, to summarise some elementary properties of (2).

#### 3.1. Behaviour of the model

An epidemic model such as (2), as compared to endemic models with demographic components, only has one possible long term outcome: a disease-free equilibrium  $E_0$  in which the only two potentially positive components are  $S$  and  $R$ . As is customary with such models, we denote  $S_\infty = \lim_{t \rightarrow \infty} S(t)$  and  $R_\infty = \lim_{t \rightarrow \infty} R(t)$  the limiting values of  $S$  and  $R$ . Because of the structure of the model, these limits always exist [3, Theorem 5.1].

As with other simple epidemic models of this type, the main interest is to know whether, following the introduction of infected individuals in the population, the number of infected individuals goes through an exponential growth phase, indicating an epidemic phase, before the disease becomes extinct. This is decided using the basic reproduction number  $\mathcal{R}_0$ . To find its value, we use the method in (Arino et al., 2007), using the notation therein. First,  $D = 1$  since  $S \in \mathbb{R}^m$  with  $m = 1$ . The matrix  $\Pi = [p_{ij}]$  (to avoid confusion with the fraction  $\pi$  of asymptomatic cases) has entry  $p_{ij}$  the fraction of the  $j$ th susceptible compartment moving, upon infection, to the  $i$ th infected compartment. Therefore, here it is a vector,  $\Pi = (1, 0, 0, 0, 0, 0)^T$ , since all new infections move to the  $L_1$  compartment. The row vector  $b$  describes the relative horizontal transmissions. It takes the form  $b = (0, \eta, 1, 1, \xi, \xi)$ . The function denoted  $\beta(x, y, z)$  in (Arino et al., 2007) is  $f(x)$  here. Finally, the matrix  $V$  describing transitions between and out of infected states takes the form

$$V = \begin{pmatrix} \varepsilon & 0 & 0 & 0 & 0 & 0 \\ -\varepsilon & \varepsilon & 0 & 0 & 0 & 0 \\ 0 & -(1 - \pi)\varepsilon & \gamma & 0 & 0 & 0 \\ 0 & 0 & -\gamma & \gamma & 0 & 0 \\ 0 & -\pi\varepsilon & 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & 0 & -\gamma & \gamma \end{pmatrix}$$

and has inverse

$$V^{-1} = \begin{pmatrix} 1/\varepsilon & 0 & 0 & 0 & 0 & 0 \\ 1/\varepsilon & 1/\varepsilon & 0 & 0 & 0 & 0 \\ (1-\pi)/\gamma & (1-\pi)/\gamma & 1/\gamma & 0 & 0 & 0 \\ (1-\pi)/\gamma & (1-\pi)/\gamma & 1/\gamma & 1/\gamma & 0 & 0 \\ \pi/\gamma & \pi/\gamma & 0 & 0 & 1/\gamma & 0 \\ \pi/\gamma & \pi/\gamma & 0 & 0 & 1/\gamma & 1/\gamma \end{pmatrix}.$$

From (Arino et al., 2007), the basic reproduction number of (2) is then

$$\mathcal{R}_0 = f(\mathbf{E}_0) b V^{-1} \Pi DS(0),$$

where  $S(0)$  is the susceptible population at the initial time. Note that for this formula to hold,  $f(\mathbf{E}_0)$  must be defined. In other words, the basic reproduction number takes the form

$$\mathcal{R}_0^{(f)} = f(\mathbf{E}_0) \left( 2 \frac{\pi \xi}{\gamma} + 2 \frac{1-\pi}{\gamma} + \frac{\eta}{\varepsilon} \right) S(0), \quad (3)$$

where the superscript  $(f)$  is used to show dependence on the nature of  $f$ . If  $f(\mathbf{x}) = \beta$ , i.e., we use mass action incidence, then

$$\mathcal{R}_0^{(\text{MA})} = \beta \left( 2 \frac{\pi \xi}{\gamma} + 2 \frac{1-\pi}{\gamma} + \frac{\eta}{\varepsilon} \right) S(0), \quad (4)$$

whereas in the case of proportional or frequency-dependent incidence,  $f(\mathbf{E}_0) = \beta/S(0)$  and

$$\mathcal{R}_0^{(\text{FD})} = \beta \left( 2 \frac{\pi \xi}{\gamma} + 2 \frac{1-\pi}{\gamma} + \frac{\eta}{\varepsilon} \right) \quad (5)$$

Observe that the latter form stems from the fact that at  $\mathbf{E}_0$ ,  $\langle \mathbf{1}, \mathbf{E}_0 \rangle = S_\infty + R_\infty$  and the latter equals  $S(0)$  since the model clearly preserves the total population.

### 3.2. Final size relations

One measure of particular importance in the context of COVID-19 and other emerging or re-emerging pathogens is the epidemic *final size*  $S(0) - S_\infty$ . The final size is often expressed in terms of the *attack rate*  $(S(0) - S_\infty)/S(0)$ .

In the case where incidence is mass action, since we have a single susceptible class, the method in (Arino et al., 2007) provides an explicit final size relation,

$$\ln \left( \frac{S(0)}{S_\infty} \right) = \frac{\mathcal{R}_0^{(\text{MA})}}{S(0)} (S(0) - S_\infty) + \beta b V^{-1} \mathbf{I}(0),$$

where  $\mathbf{I}(0) = (L_1(0), L_2(0), I_1(0), I_2(0), A_1(0), A_2(0))^T$  is the initial infected population. This simplifies to

$$\ln \left( \frac{S(0)}{S_\infty} \right) = \frac{\mathcal{R}_0^{(\text{MA})}}{S(0)} (S(0) - S_\infty + L_1(0) + L_2(0)) + \frac{\beta}{\gamma} (2I_1(0) + I_2(0) + 2\xi A_1(0) + \xi A_2(0)) \quad (6)$$

With other incidence functions, we obtain inequalities of the form

$$\ln \left( \frac{S(0)}{S_\infty} \right) \geq \frac{\mathcal{R}_0^{(f)}}{S(0)} (S(0) - S_\infty) + f(K) b V^{-1} \mathbf{I}(0),$$

where  $K$  is the initial total population size. Similarly to (6), this can easily be simplified but is not shown here.

Thus, in the case of mass action incidence, the most used incidence in the case of epidemics, finding the final size of the epidemic (or, equivalently, the epidemic attack rate) requires to solve the simple transcendental equation (6), which is easily done at least numerically.

### 3.3. Detecting the turning point and the peak

Other measures of particular importance in the context of the study of an epidemic outbreak are the timing of the peak as well as the disease prevalence at the peak. In the discussion that follows, assume incidence is mass action.

In a classic ODE Kermack-McKendrick SIR model, the peak is easily characterised as the point in phase space where  $I' = 0$ . Because of the number of compartments, the peak here can be studied in two parts. Before the peak is actually reached, there

is first a point in time at which incidence starts to decrease; this is sometimes called the *turning point* of the epidemic. In the present model, this point is reached when  $L_1' = 0$ , i.e., with mass action incidence, when

$$S = \frac{\varepsilon L_1}{\beta(\eta L_2 + I_1 + I_2 + \xi(A_1 + A_2))} =: \Psi(\mathbf{I}), \quad (7)$$

where  $\mathbf{I} = (L_1, L_2, I_1, I_2, A_1, A_2)$ . Note that  $\Psi$  is not defined everywhere in phase space; for instance, it is not defined at  $\mathbf{E}_0$  nor if one considers an  $\mathbf{I}$  with sign pattern  $(+, 0, 0, 0, 0, 0)$ . However, past initial transients and before  $\mathbf{I}$  becomes close to  $\mathbf{E}_0$ , (7) provides a characterisation of the phase the epidemic is currently in. If  $S > \Psi(\mathbf{I})$ , then the “natural tendency” of the epidemic is to propagate more. When enough susceptible individuals have been “consumed” by the infection, i.e., when  $S < \Psi(\mathbf{I})$ , the epidemic cannot sustain itself anymore. For perspective, in the context of COVID-19 and using time units of days and mass action incidence,  $\varepsilon$  is of the order of  $10^{-1}$  while  $\beta$  is typically several order of magnitude smaller, so the factor  $\varepsilon/\beta$  in (7) is in the range  $[10^4, 10^7]$ . Note that because only  $I_1$  and  $I_2$  are observable, determining whether (7) holds is impossible in the field. On the other hand, this is easily done when considering numerical simulations.

The *peak* is then the point at which prevalence of the infection in the population is maximum. Here, we focus on the *observable peak*, i.e., the one that can actually be measured in real life data. Because of that, we consider the peak to happen at the time when the observable part of the epidemic, i.e., the number of infectious individuals  $I_1 + I_2$ , is maximum. Thus, at the peak,  $I_1' + I_2' = 0$ , i.e.,

$$(1 - \pi)\varepsilon L_2 = \gamma I_2. \quad (8)$$

Since  $(1 - \pi)\varepsilon L_2$  is the rate at which new symptomatic infections occur and  $\gamma I_2$  is the rate at which symptomatic infections are resolved, either by recovery or death, the occurrence of conditions necessary for the peak to take place can be inferred from the data. So, while (7) is not observable, (8) is.

### 3.4. Start date of the epidemic in a location

When conducting a numerical investigation of the system properties, finding the date at which to start simulations is important. We describe here the methodology used to do so.

In this work, we do not account for “structural” under-reporting of cases and thus assume that the observable quantity in our model, in terms of infection, is the number of new symptomatic cases. As a consequence, we take, for a given location, the time  $t_c$  at which the first  $c$  confirmed cases are reported. A location is thus characterised by a pair  $(t_c, c)$ .

For illustration, take the situation of cumulative confirmed case counts in China as reported to WHO, which was of 547 cases on January 22, 2020, i.e.,  $(t_c = 2020\_01\_22, c = 547)$ . We seek the initial date  $t_i$  for China such that, on January 22, 2020, China has this cumulative confirmed case count when solving (2) numerically and with the parameters considered. Given a point  $u$  in parameter space, we initiate a simulation with initial time  $t_0 = 0$ . We solve (2) numerically, forward in time over the interval  $[0, t]$ , with  $S(0)$  the population of China,  $L_1(0) = 1$  and all other state variables equal to zero. This gives a solution  $\mathbf{x}(t, t_0 = 0, u)$ . Extracting  $L_2(t, t_0 = 0, u)$  from this solution, we compute

$$C(t) = \int_{t_0=0}^t (1 - \pi)\varepsilon L_2(s, t_0, u) ds.$$

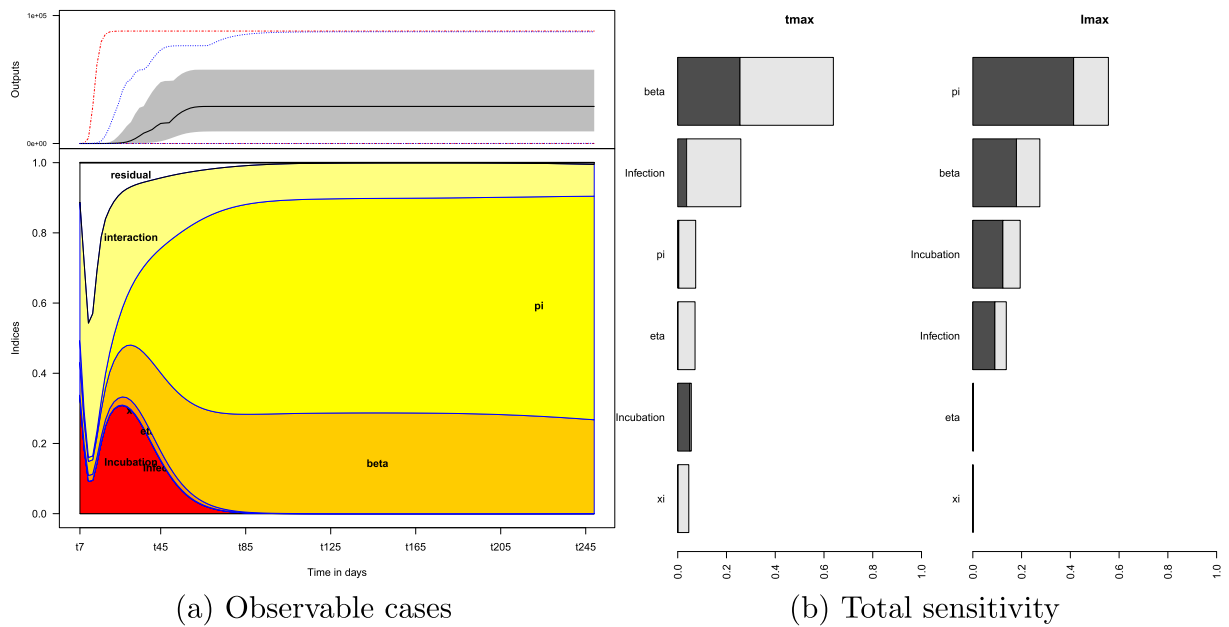
As  $(1 - \pi)\varepsilon L_2$  is the rate at which individuals enter the  $I_1$  compartment,  $C(t)$  represents the total number of individuals having become infectious at time  $t$ , i.e., the cumulative number of symptomatic infections  $t$  days after the introduction of the first infectious individual in the population. We then let  $t^*$  be the point when  $C(t^*) = 547$ ;  $t_i$  for China, with the parameters  $u$ , is then  $t_i = 2020\_01\_22 - t^*$ .

Note that a random point in parameter space might lead to a situation where  $\mathcal{R}_0 < 1$ . This is used in parameter estimation to disqualify such points. Other points that would be disqualified are those such that  $t^*$  is too large. These aspects will be discussed in further work on the topic.

**Table 1**

Model parameters. Incubation time and Infectious period give the parameters  $\varepsilon = 2/(\text{Incubation time})$  and  $\gamma = 2/(\text{Infectious period})$ , respectively.

Parameter	Definition	Range
$\beta$	transmission coefficient	$[5 \times 10^{-7}, 10^{-4}]$
$\eta$	attenuation of transmission for incubating individuals	$[0, 0.2]$
$\xi$	attenuation of transmission for asymptomatic individuals	$[0, 1]$
$\pi$	fraction of asymptomatic cases	$[0, 1]$
Incubation time	mean duration of incubation	$[1, 14]$
Infectious period	mean duration of infectious period	$[2, 14]$



**Fig. 3.** Sensitivity analysis results: (a) Effects of factors on the number of observable cases as a function of time. The upper panel shows inter-quartile (grey area) and median (bold line) output values over time. The lower panel displays the sensitivity indices over time for the main effects and the first-order interactions (interaction between two factors). Residuals correspond to contribution to the variance from interactions between three or more factors. (b) Effects of factors on the time of the peak  $t_{\max}$  and size of the peak  $I_{\max}$ . Bars represent total indices of sensitivity for each factor. The dark grey parts of bars are the main effect sensitivity indices, light grey parts correspond to first-order interactions.

#### 4. Sensitivity analysis

A global sensitivity analysis is carried out to characterise the impact of uncertainty of factors (inputs of the model) on model outputs. The inputs of model and their ranges are listed in Table 1;  $p$  inputs associated to the model parameters are considered. The number of observable cases during the course of the epidemic and at the peak, and the timing of the peak are the outputs of interest here. The number of observable cases is the cumulative confirmed case count  $C(t)$ .

The variance-based analysis is performed using the R package *multisensi*. For each factor,  $m = 7$  values are chosen using Latin hypercube sampling with uniform distributions on the range considered for the factor of interest (Table 1). Then, using a complete factorial design approach,  $m^p$  scenarios are generated. Simulations are run for these  $m^p = 7^6 = 117,649$  scenarios with the same initial condition  $S(0) = 100,000$ ,  $I_1(0) = 1$  and  $L_1(0) = L_2(0) = I_2(0) = R(0) = 0$ . Then, sensitivity indices are computed using a classic ANOVA decomposition (Lamboni, Makowski, Simon, Gabrielle, & Monod, 2009; Monod, Naud, & Makowski, 2006).

Fig. 3a shows the influence of factors on the number of observable cases over time. The sensitivity indices are computed every 2 days over a period of 250 days. The lower panel of Fig. 3a details, at each time point, contributions (normalised to 1) of model parameters to the total variability of model responses. At a given time  $t_j$ , the relative lengths of the coloured segments represent the relative contributions of the main effect sensitivity indices to this total variability. Interactions between two factors is denoted “interaction” and those involving three or more factors is denoted “residual”.

For instance, at  $t_{85}$  (after 85 days), the number of observable cases is mostly sensitive to the main effect of  $\pi$ . The main effect of the transmission coefficient  $\beta$  accounts for one-third of the variability. Interactions between factors contribute to about one-tenth of the output variability. In the long-run, this trend persists.

In the early dynamics, the variability is due to the main effects of incubation time and interactions between factors.

Fig. 3b displays the total sensitivity, a measure of influence of each factor; the outputs considered here are the timing of the peak and value of the observable variables at the peak,  $t_{\max}$  and  $I_{\max}$  respectively. In each bar (total sensitivity), the main effect (dark grey) and interaction between the factor of interest and another (light grey) are detailed. The time of the peak  $t_{\max}$  is affected by the transmission  $\beta$  and incubation time. The value of the peak  $I_{\max}$  is mainly influenced by the proportion of asymptomatic  $\pi$ , the duration of incubation and infectiousness and transmission  $\beta$ . The transmission coefficient ( $\beta$ ) has the most influence on  $t_{\max}$  whereas  $I_{\max}$  is most impacted by the proportion  $\pi$  of asymptomatic cases. Furthermore, note that for the most influential factors of  $t_{\max}$  the first-order interactions contribute the most to their total sensitivity. The timing of peak of observable case numbers results mostly from the interplay between factors.

## 5. Discussion

The model presented here can be used to consider some of the aspects of spread of a novel or re-emergent pathogen. We have focused here on practical aspects of the use of the model, focusing on the need, in emergency response settings, to provide a fast evaluation of outcomes.

Model (2) consists of 8 differential equations, but its parametrisation involves the same number of parameters as the 5 equations model from which it is derived (Arino et al., 2006). Because a lot is unknown during the early stages of a crisis like the ongoing COVID-19 pandemic, simple models that can be fitted using a minimal number of parameters are extremely useful.

In a time of crisis, it is however also important not to oversell the capabilities of a model. We have been considering many variations on the current model, as part of work conducted in Canada regarding COVID-19. While this base model has proved very helpful in many circumstances, it is unable in particular to provide insights into testing or contact tracing. Other modelling paradigms such as individual-based (IBM) or agent-based (ABM) models are much better suited to answer questions in this area. A continuous-time Markov chain model version of this model has for instance been considered to answer specific questions where a better understanding of the infection chains is required, such as importations of cases into new locations.

Where our model is quite appropriate, on the other hand, is when reasonably sized populations are considered. In this case, it can easily be shown, as had been done in (Arino et al., 2006), that ODE models provide essentially similar results to population-level IBM and ABM.

This highlights another strength of this type of approach: ODE models are quite amenable to extensive sensitivity analyses, bringing forward an important non methodological conclusion of the present work. One of the most influential factor/parameter in the model considered is the proportion of asymptomatic cases. The basic reproduction number  $\mathcal{R}_0$  (through the transmission parameter  $\beta$  and its interaction with other parameters) determines the time of the peak of observable cases whereas the value of the peak, and thus the impact on the health care system, depends critically on the proportion of cases that are asymptomatic. This highlights the imperious need for more research, both in the field and in modelling, to understand the drivers of asymptomaticity and its prevalence among cases.

## Declaration of competing interest

We have no conflict of interest.

## Acknowledgements

The authors are supported in part by NSERC Discovery Grants. JA is also supported by CIHR through the Canadian 2019 Novel Coronavirus (COVID-19) Rapid Research Funding Opportunity. The authors wish to thank Dr. Nicholas Ogden, Director of Public Health Risk Science (PHRS) at the National Microbiology Laboratory of the Public Health Agency of Canada, as well as Drs. Aamir Fazil and Erin Rees, also with PHRS, for helpful discussions during work on COVID-19.

## References

- Arino, J. (2020). Mathematical epidemiology in a data-rich world. *Infectious Disease Modelling*, 5, 161–188. <https://github.com/julien-arino/modelling-with-data-for-code>.
- Arino, J., Brauer, F., van den Driessche, P., Watmough, J., & Wu, J. (2006). Simple models for containment of a pandemic. *Journal of The Royal Society Interface*, 3(8), 453–457.
- Arino, J., Brauer, F., van den Driessche, P., Watmough, J., & Wu, J. (2007). A final size relation for epidemic models. *Mathematical Biosciences and Engineering*, 4(2), 159–175.
- Backer, J. A., Klinkenberg, D., & Wallinga, J. (2020). Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from Wuhan, China, 20–28 January 2020. *Euro Surveillance*, 25(5).
- Hu, Z., Song, C., Xu, C., Jin, G., Chen, Y., Xu, X., et al. (2020). Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. medRxiv.
- Jin, Z., Zhang, J., Song, L.-P., Sun, G.-Q., Kan, J., & Zhu, H. (2011). Modelling and analysis of influenza A (h1n1) on networks. *BMC Public Health*, 11(Suppl 1), S9.
- Kim, S., Lee, J., & Jung, E. (2017). Mathematical model of transmission dynamics and optimal control strategies for 2009 A/H1N1 influenza in the Republic of Korea. *Journal of Theoretical Biology*, 412, 74–85.
- Lamboni, M., Makowski, D., Simon, L., Gabrielle, B., & Monod, H. (2009). Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research*, 113(3), 312–320.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., et al. (03 2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*. <https://doi.org/10.7326/M20-0504>.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H., et al. (2020). The incubation period of 2019-ncov from publicly reported confirmed cases: Estimation and application. medRxiv.
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., et al. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (covid-19). medRxiv.
- Monod, H., Naud, C., & Makowski, D. (2006). Uncertainty and sensitivity analysis for crop models. *Working with dynamic crop models: Evaluation, analysis, parameterization, and applications*, 4, 55–100.
- Z.-D. Tong, A. Tang, K.-F. Li, Peng L., H.-L. Wang, J.-P. Yi, et al. Potential presymptomatic transmission of SARS-CoV-2, zhejiang province, China, 2020. *Emerging Infectious Diseases*, 26(5), 2020.