The SIcR model with time since infection and its Galerkin approximation

Joseph D. Peterson, University of Cambridge, DAMTP

June 18, 2020

Abstract

Epidemic models are useful tools in the fight against infectious diseases. However, the need for fine-grained resolution of details related to geography and demographics often constrains the amount of detail that one can include in the underlying model for disease transmission. At the same time, simplifications to the model of disease transmission will propagate through the rest of the epidemic model with effects that are generally difficult to anticipate. Therefore, a good disease transmission model must strike a balance between accuracy and complexity. Here, we consider an extension of the classic SIR description of disease transmission that accounts for an infected class whose infectiveness varies with time since infection. We refer to this model as the SIcR model, where 'c' stands for a 'continuous' resolution of time-since-infection. These equations have been previously described, but here we show that numerical solutions of the SIcR model can be solved via a Galerkin approximation for which the structure of the governing equations is compatible with existing empidemic modelling framworks based on the SIR model. Finally, we share some predictions for the time-since-infection model and compare against predictions of the standard SIR model.

1 Background and Introduction

Epidemic modelling is a valuable tool for making well-controlled predictions about how a disease is likely to spread through a population. Predictions from these models can inform strategies that mitigate the damage said disease will cause. Better strategies are possible with better epidemic modelling, and ultimately an epidemic model is only as reliable as the assumptions it makes.

Good epidemic models must be able to provide fine-grained descriptions for how a disease is spreading. It is not enough to simply predict a number of infected persons – one must also be able to predict the spread of disease from city-to-city and age group to age group, for example [1, 2, 3, 4]. Given the burden of complexity an epidemic model is expected to bear, one cannot usually afford much more than an elementary model of disease transmission at each level.

To that end, some of the most popular classes of epidemic models are generalizations of the standard SIR framework, in which a population is partitioned into susceptibles (S), infecteds (I), and recovered (R) [5]. In this framework, a susceptible individual becomes infected by encountering an infected person in some way or another. An infected person is immediately infectious, and the subsequent transition to recovered is a random process following a Poisson distribution. The SIR model does capture some important features of disease transmission, but its description of the infected population is inflexible and generally at odds with the medical realities of how diseases progress.

A more realistic model of disease transmission will model the infectiousness of an infected individual as a function of the time since infection [5, 2, 6] Does a person become infectious right away, or is there a delay of several days? When is a person most infectious? After how long can a person be considered recovered? The answer to these questions will shape how a disease spreads, and so epidemic models that track time since infection in the infectious population are very important [7, 8].

On the surface, it might seem that any epidemic model based on time since infection will be substantially more complex than the basic SIR model and therefore of limited value to fine-grained epidemic modelling applications. In this report, however, we will show that the SIR model can be extended to consider time-since-infection without making the model much more difficult to solve.

2 Governing Equations

2.1 The standard SIR model

First, we review the standard SIR model. The number of susceptibles, S, infecteds, I, and recovereds, R evolve in time by:

$$\frac{dS}{dt} = -\frac{\beta}{N} IS \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta}{N} IS - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

where β and γ are rate constants for infection and recovery, respectively (units of 1/time) and $N = S + I + R$ is the total population.

2.2 Extension to time since infection (TSI)

Models that include time since infection are not new, and in fact the equations presented here are covered by existing models for epidemics, many of which are more general [2, 6, 9]. However, models with time-since-infection have historically been computationally expensive. As a compromise between accuracy and usability, 'multi-stage' models are a popular strategy for approximating true time-since-infection models [10, 8]. Here, however, we will show that this compromise is not necessary – with the right numerical methods, one can obtain all the accuracy of time-since-infection for the same computational cost as a typical multi-stage model.

In a model with time since infection, we will speak of the infected class in terms of the number density of persons (per unit time) whose infections began at a time s prior to the current time t, $I(t,s)$. Thus, the total number of persons infected in the narrow interval between s and $s+\delta s$ prior to time t is given by $I(t,s)\delta s$. As time passes, the time since infection for all infected persons also continues to pass in the same way. Given $I(t,s)$, one can calculate the rate at which susceptible persons become infected:

$$\frac{dS}{dt} = -\int_0^T ds \frac{\beta(s)}{N} I(t,s)S \tag{4}$$

Note that the mean rate constant for infection $\beta(s)$ now varies with the time since infection. Here we have assumed that infections older than time T are no longer transmitting, $\beta(s > T) = 0$, which allows the integral to be truncated at time T. We have also ignored the possibility of re-infection for the time being. The number density equation for the infected class evolves in time due to the passage of time,

$$\frac{\partial}{\partial t} I(t,s) + \frac{\partial}{\partial s} I(t,s) = 0 \tag{5}$$

and the number density of new infections must match the rate at which susceptible individuals are becoming infected:

$$I(t,0) = \int_0^T ds \frac{\beta(s)}{N} I(t,s)S \tag{6}$$

As with the standard SIR model, the equations for susceptibles and infecteds are closed. However, for pracitical applications of an epidemic model one may also need to describe the various sub-classes of the infected population (e.g. infectious, recovered, hospitalized, deceased, quarantined, asymptomatic, etc), with each sub-class α representing a fraction $\Phi^\alpha(s)$ of the total with number density$I^\alpha(t,s) = \Phi^\alpha(s)I(t,s)$. The total population in each sub-class at any time is then given by:

$$I^\alpha(t) = \int_0^\infty ds \Phi^\alpha(s) I(t,s) \tag{7}$$

This partitioning is also represented graphically in Figure 1.

Note also that if each sub-class has a rate constant for infection $\beta_\alpha(s)$, then the mean value $\beta(s)$ can be pre-computed by:

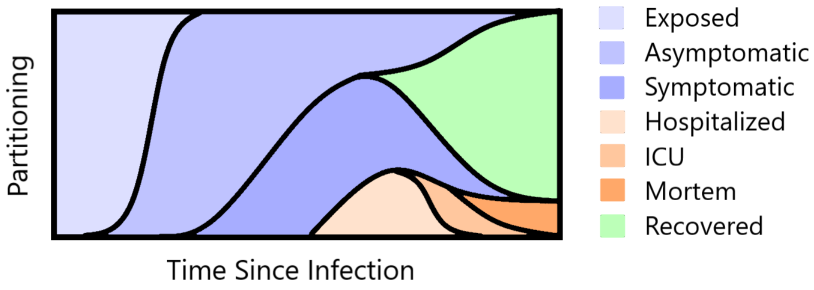

Figure 1: A graphical representation of how an infected population might be partitioned into different categories.

$$\beta(s) = \sum_{\alpha} \Phi^{\alpha}(s)\beta_{\alpha}(s) \tag{8}$$

For the analysis in this report, we will only present calculations for two subclasses of infecteds – those who are within the infectious range and those who are not. We will refer to the latter group as ’recovered’ with population R, and $\Phi_R(s)$ is just a step function at $s = T$. Taking a time derivative of equation 7, this can also be written as:

$$\frac{dR}{dt} = I(t, T) \tag{9}$$

Or, more generally, given $\phi^{\alpha}(s) = d\Phi^{\alpha}/ds$ the population in any subclass of infecteds evolves by:

$$\frac{dI^{\alpha}}{dt} = \int_0^T ds \phi^{\alpha}(s) I(t, s) \tag{10}$$

In general, one can choose any number of subclasses with arbitrarily complex shapes for each $\Phi_{\alpha}(s)$ with minimal added cost for obtaining solutions numerically (provided one chooses an appropriate numerical method). We refer to equations 4 - 7 as the SIcR model, where ‘c’ stands for a continuous resolution of the time since infection.

2.3 Additional Remarks on Partitioning the Infecteds

In equations 4 - 7, it was assumed that the functions Φ^{α} partitioning the infecteds depend on time since infection but not on time itself. This approximation works well for modelling subclasses whose populations reflect the nature of the disease itself (like hospitalizations and fatalities), but it is less suited to subclasses whose populations reflect a behavioral or societal response to the disease. Consider a system in which quarrantines are assigned by a track-and-trace progam – the probability of being under quarantine at time since infection s depends on the history of testing to that point and not necessarily on the current state of testing. Therefore, in the most general case on can write a dynamical equation:

$$\frac{\partial \Phi^{\alpha}}{\partial t} + \frac{\partial \Phi^{\alpha}}{\partial s} = \phi^{\alpha}_{in}(t,s) - \phi^{\alpha}_{out}(t,s) \tag{11}$$

where $\phi^{\alpha}_{in}(t,s)$ and $\phi^{\alpha}_{out}(t,s)$ are the probability density for moving in/out subclass α at time t and time since infection s. Note, however, that if the criterion for moving people in and out of subclass α are changing on timescales that are slow compared to the typical residence time for subclass α, then to leading order one can apply a quasi-static approximation for which equation 11reverts to:

$$\Phi^{\alpha}(t,s) = \int_0^s ds' \phi^{\alpha}(t,s') \tag{12}$$

$$\phi^{\alpha}(t,s) = \phi^{\alpha}_{in}(t,s) - \phi^{\alpha}_{out}(t,s) \tag{13}$$

Note that in this case the population in each subclass of infecteds evolves as:

$$\frac{dI^{\alpha}}{dt} = \int_0^T ds \phi^{\alpha}(t,s) I(t,s) \tag{14}$$

Finally, while the underlying probability density functions $\phi^{\alpha}(t,s)$ can be obtained from track-and-trace data, we have found that the available data is not sufficiently detailed to precisely specify all necessary inputs to a time since infection model. To that end, we suggest that the probability density functions can be approximated as beta distributions:

$$\phi^{\alpha}_{in}(s) = p^{\alpha}_{in} \frac{1}{T} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{s}{T}\right)^{1-a} \left(1 - \frac{s}{T}\right)^{1-b} \tag{15}$$

where p^{α}_{in} is the probability that an infected person is ever assigned to subclass α and the 'shape' of the distirubiton is defined by parameters a and b. The same parameterization can be done for each ϕ^{α}_{out} as well. Finally, the shape parameters α and β can be explicitly related to the mean $\bar{x}$ and variance v of the distribution:

$$a = \frac{\bar{x}^2(1-\bar{x})}{v} - \bar{x} \tag{16}$$

$$b = \left(\frac{\bar{x}(1-\bar{x})}{v} - 1\right)(1-\bar{x}) \tag{17}$$

Given preliminary guesses for the mean and variance of entry/exit times to each subclass of infecteds, one can fully specify the time since infection model and then use standard inference and optimization tools to obtain improved estimates and uncertainties. This same parameterization strategy also applies to the rate constant for infection, $\beta(s)$, which is itself a probability density function for transmitting the disease at time since infection s.

2.4 Nondimensionalization

Before proceeding further, it is helpful to recast the governing equations into dimensionless form. We rescale S and R by the total population N, such that $\tilde{S} = S/N$ and $\tilde{R} = R/N$. Using a characteristic time $T_C = T/2$, we rescale time as $\tilde{t} = t/T_C$ and number density of infecteds as $\tilde{I} = IT_C/N$. The time since infection is rescaled to the variable $\tilde{s} = s/T_C - 1$ so that the range $s \in [0, T]$ is mapped to $\tilde{s} \in [-1, 1]$. Finally, given the reproduction number $R_0 = \int_0^T \beta(s)ds$ we rescale$\tilde{\beta} = \beta T_C/R_0$ so that $\int_{-1}^{1} d\tilde{s}\tilde{\beta}(\tilde{s}) = 1$.

For compactness of notation, we will suppress tildes in our dimensionless variables – from this point forward, all variables are dimensionless unless otherwise stated. The dimensionless governing equations of the SIcR model with time since infection are given by:

$$\frac{dS}{dt} = -R_0 S \int_{-1}^{1} ds\beta(s)I(t,s) \tag{18}$$

$$\frac{\partial}{\partial t}I(t,s) + \frac{\partial}{\partial s}I(t,s) = 0 \tag{19}$$

$$I(t,-1) = R_0 S \int_{-1}^{1} ds\beta(s)I(t,s) \tag{20}$$

$$\frac{dR}{dt} = I(t,1) \tag{21}$$

These equations can be solved by any number of means. In the section that follows, we will present three discretization stategies for numerical solution. First, we show that existing multi-stage models can be derived using a particularly inefficient method-of-lines discretization. Second, we show that with a few small changes the accuracy of the standard multi-stage model can be improved dramatically at a small cost – time-stepping is inflexible. Finally, for good accuracy and flexible time-stepping, we present a novel discretization based on a Galerkin approximation. The Galerkin approximation is promising in principle, but in practice it is less robust whenever dynamics are non-smooth or unsteady.

2.5 Method of Lines: the SIkR model

The PDE can be converted to an approximate ODE representation via the method of lines. Here, we discretize s in uniform intervals of width h and represent advection/integration in s via upwinding and right Riemann sums, respectively. These methods are robust but generally poor performing, with accuracy $O(h)$ -- about as bad as any stable numerical scheme can be. We will show that this choice of discretization maps the SIcR to the standard multi-stage generalization of SIR model [10]. Details about the truncation errors introduced by this particular choice of discretization scheme will be discussed shortly.

We evaluate the solution at discrete points in s, evaluating at $s_1, s_2, \ldots, s_N$ where $s_n = -1 + (n-1)h$ and $h = 2/(N-1)$. In this section, we denote I_n as the number density of infected persons at each s_n and $\beta_n = h\beta(s_n)$. Applying this discretization scheme to the PDE equations, we obtain evolution equations for S, I_n, and R:

$$\frac{dS}{dt} = -R_0 S \sum_{n=2}^{N} \beta_n I_n \tag{22}$$

$$I_1 = R_0 S \sum_{n=2}^{N} \beta_n I_n \tag{23}$$

$$\frac{dI_n}{dt} = \frac{1}{h}(I_{n-1} - I_n) \tag{24}$$

$$\frac{dR}{dt} = I_N \tag{25}$$

This maps exactly to the SIR model when $N = 2$ or to a multi-stage SIkR model with $k = N - 1$ stages more generally. Therefore, we confirm that in the limit of $h \to 0$, the SIkR model is equivalent to a true time-since infection model. In practice, however, the convergence is so slow that truncation errors will have a large influence on the predictions. Therefore, we discuss the truncation errors introduced by this scheme and how they systematically influence the resulting predictions.

First, we consider truncation error when integrating over s using left Riemann sum. Assuming $\beta_n = \beta_0$ for all n, the right Riemann sum will under/over estimate the infectivity of the infected class during the rising/falling of the epidemic by $O(h)$. Next, upwinding to approximate derivatives in s introduces numerical diffusion, meaning that the numerical solution behaves like a PDE model to which an additional diffusion term has been added:

$$\frac{\partial}{\partial t} I(t,s) + \frac{\partial}{\partial s} I(t,s) = D \frac{\partial^2}{\partial s^2} I(t,s) \tag{26}$$

Where the numerical diffusion coefficient D scales as $D \sim O(h)$. Thus, the upwinding scheme tends to smear out distinctions in the age-profile of infecteds. This error is less important if $\beta(s)$ is constant, but in general the resulting numerical diffusion will tend to over-estimate the number of late/early stage infected during the rising/falling of the epidemic by $O(h)$. Numerical diffusion can be eliminated if the discretized ODE problem is integrated via forward Euler with a time-step of exactly h – this moves the truncation error from the $s-$domain into the $t-$domain, but it keeps the overall error at $O(h)$ and allows for faster calculation.

2.6 Method of Lines: Predictor/Corrector

In the preceding section, we showed that the SIkR model is an approximation of a TSI model with first order accuracy, $O(h)$. The accuracy can be improved to second order $O(h^2)$ with a few small modifications. We use a trapezoid rule (as opposed to a Riemann sum) for quadrature and a predictor-corrector midpoint rule for evolving the susceptible and recovered populations. The infected population (with the exception of newly infecteds) are upwinded with a Courant number of one to eliminate numerical diffusion. The principle disadvantage of this method is that time-stepping is inflexible.

We evaluate the solution at discrete points in s, evaluating at $s_1, s_2, s_3 \ldots, s_N$ where $s_n = -1+(n-1)h$ and $h = 2/(N-1)$ and also at discrete times $t^{(k)} = kh$ for integer k. We denote I_n as the number density of infected persons at each s_n and $\beta_n = h\beta(s_n)$ for $n \neq 0, N$ and $\beta_n = h\beta(s_n)/2$ for $n = 0, N$. The time-step $k = 0, 1, 2, ...$ will be indicated as a superscript in parentheticals, e.g. $S^{(k)}$is the susceptible population at time $t^{(k)}$. In the language of a predictor/corrector method, intermediate predictions will be further denoted by the letter p inside the parenthetical, e.g. $S^{(k+1,p)}$.

Beginning from time $t^{(k)}$, we can calculate the state of the system at time $t^{(k+1)} = t^{(k)} + h$ as follows. We begin with an explicit prediction step:

$$\Delta S^{(k)} = h\left[-R_0 S^{(k)} \sum_{n=1}^{N} \beta_n I_n^{(k)}\right] \tag{27}$$

$$S^{(k+1,p)} = S^{(k)} + \Delta S^{(k)} \tag{28}$$

$$I_0^{(k+1,p)} = -\Delta S^{(k)}/h \tag{29}$$

$$I_{n>0}^{(k+1,p)} = I_{n-1}^{(k)} \tag{30}$$

Then, we proceed to the correction:

$$\Delta S^{(k+1,p)} = h\left[-R_0 S^{(k+1,p)} \sum_{n=1}^{N} \beta_n I_n^{(k+1,p)}\right] \tag{31}$$

$$S^{(k+1,p)} = S^{(k)} + \frac{1}{2}\Big(\Delta S^{(k)} + \Delta S^{(k+1,p)}\Big) \tag{32}$$

$$I_0^{(k+1)} = -\Delta S^{(k+1,p)}/h \tag{33}$$

$$I_{n>0}^{(k+1)} = I_{n-1}^{(k+1,p)} \tag{34}$$

$$R^{(k+1)} = R^{(k)} + \frac{h}{2}\Big(I_N^{(k)} + I_N^{(k+1)}\Big) \tag{35}$$

This method has second order accuracy $O(h^2)$, so compared to the SIkR model one needs only a few stages to obtain good results. The principle weakness of this method is that time-stepping is inflexible – each time step must advance the solution forward by a time of exactly h. Note also that for numerical stability, the time-step h must be chosen to be sufficiently small (e.g. $h(R_0 - 1){<}1$).

2.7 Expansion in Legendre Polynomials (Galerkin)

For spectral accuracy in s and flexible time-stepping, we provide a discretization based on a Galerkin approximation. On the interval $s \in [-1, 1]$, we can write $I(t, s)$ as weighted sum of Legendre polynomials:

$$I(t,s) = \sum_{n=0}^{\infty} I_n(t) P_n(s) \tag{36}$$

Any set of orthogonal basis functions could be used for this purpose, but we have found that Legendre polynomials work particularly well in this application. If we truncate the sum after $m{+}1$ terms and insert into the governing equations, we obtain a closed system of equations for $S, I_0, I_1, I_2, \ldots, I_m$, and R. This technique is known as a Galerkin approximation.

The population of susceptibles evolves by:

$$\frac{dS}{dt} = -R_0 S \sum_{n=0}^{m} a_n I_n \tag{37}$$

where the coefficients $a_n$are:

$$a_n = \int_{-1}^{1} ds \beta(s) P_n(s) \tag{38}$$

For $n < m$, we obtain evolution equations for the coefficients I_n:

$$\frac{dI_n}{dt} + \sum_{k=0}^{m} b_{nk} I_k = 0 \tag{39}$$

$$b_{nk} = \left[\frac{2n+1}{2}\right] \int_{-1}^{1} ds P_n(s) P_k'(s) \tag{40}$$

The coefficients b_{nk} evaluate to $b_{nk} = 2n + 1$ if $n + k$ is odd and $k > n$, otherwise they evaluate to zero.

The highest order polynomial coefficient I_m is constrained by the boundary condition at s=0 (c.f. equation 20):

$$\sum_{n=0}^{m} I_n (-1)^n = R_0 S \sum_{n=0}^{m} a_n I_n \tag{41}$$

In practice, this constraint can be used to eliminate the highest order Legendre polynomial coefficient, I_m, as an independent variable.

$$I_m = [(-1)^m - R_0 S a_m]^{-1} \left[S R_0 \sum_{n=0}^{m-1} a_n I_n - \sum_{n=0}^{m-1} I_n (-1)^n \right] \tag{42}$$

However, this trick of eliminating I_m only works when the reproduction number R_0 (or more generally, the contact matrix in an age-structured model) is not varying in time. In a later subsection, we will show that the boundary constraint must be handled implicitly to preserve spectral accuracy in the method. Implicit algebraic constraints are typically costly when performing numerical integration, so the predictor/corrector method may be preferred in these cases.

Another notable drawback of the Galerkin discretization is that it is ill-suited to describing non-smooth dynamics. There are two notable circumstances in which non-smooth dynamics might become important in the context of a TSI model. First, simulating an abrupt lockdown at t_L will create a discontinuity in $I(t,s)$ at $t-(s+1)=t_L$. Second, an arbitrary choice of initial condition for $I(t,s)$ is likely to create discontinuities in higher order derivatives of $I(t,s)$ at $t-(s+1)=0$. When discontinuities are likely to be present in the solution, it may be preferable to use the method of lines discretization until the discontinuity ages to $s>1$and then switch back to the Galerkin discretization.

3 Sample Results

3.1 Linear Growth

Here, we return to the PDE formulation of the model to consider linear growth or decay of a disease when infectious cases are present in small numbers. From equation 19 we find:

$$I(t,s) = I_0 \exp(\lambda(t-s)) \tag{43}$$

with growth rate λ and density of recent infecteds $I_0 \ll 1$. From the boundary condition at $s=-1$ we also have:

$$I(t,-1) = I_0 \exp(\lambda(t+1)) \tag{44}$$

This gives an implicit relationship between the growth rate λ and the reproduction number R_0 and $\beta(s)$ that can be solved numerically:

$$\frac{1}{R_0 S} = \int_{-1}^{1} \beta(s) \exp(-\lambda(s+1)) ds \tag{45}$$

Thus, we see that the growth rate λ (in units of $2/T$) is not uniquely determined by the reproduction number R_0 but also depends on how the infectiousness of a disease rises and falls over time within the infectious period.

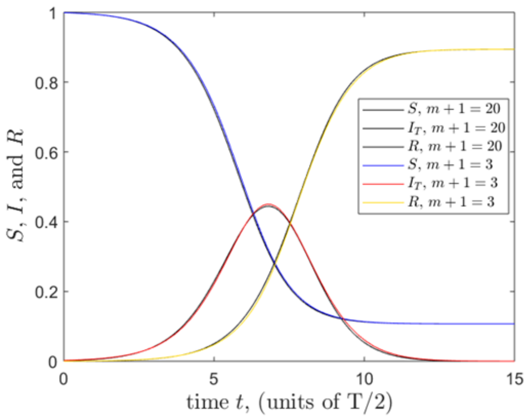

Figure 2: Comparing results for the time-since-infection SIR model with $m+1 = 3$ and $m+1 = 20$ Legendre polynomials. The results are essentially indistinguishable, meaning that accurate predictions of the time-since-infection model are not much more costly than the standard SIR model.

3.2 Nonlinear Dynamics

Using $R_0 = 2.5$ and $\beta(s) \sim (1+s)^2(1-s)^2$ and an initial condition of $I(0,s) = 10^{-3}\exp(-\lambda s)$, we obtain predictions for the fraction of a population that is classified as susceptible, infected, and recovered. These results are presented as proof-of-concept only and do not purport to describe the progression of any particular disease.

First, how many Legendre polynomials are required for predictions of the model to converge? In Figure 2, results are presented for $m+1 = 3$ Legendre polynomials and $m+1 = 20$ Legendre polynomials. It is clear to see that there is very little difference between the two, and the expansion of a single infected class into three subclasses comes with a negligible computational cost. This quality of convergence does not generally depend on any constraint of smoothness in the function $\beta(s)$, but more Legendre polynomials will be needed for diseases that are more infectious (larger R_0). Here, we have defined the fraction of the population that is infectious at time t by $I_T(t) = \int_{-1}^{1} ds I(t,s)$.

Next, how does this picture of an epidemic's progression differ from that provided by an SIR model with the same R_0? In Figure 3, we find that the two models do show good agreement for the final state of the system, but the transients differ considerably. The differences in transients can be partly explained by different choices for rescaling the time domain, but the SIR model also predicts a lower peak to the epidemic and a more rapid initial growth of the recovered class. These qualitative differences can be attributed to the distinct stages of infectiousness detailed in the TSI model – in particular, the extended periods of low infectiousness broaden the window of time for which a person is considered infectious. This leads to a higher peak number of infected persons

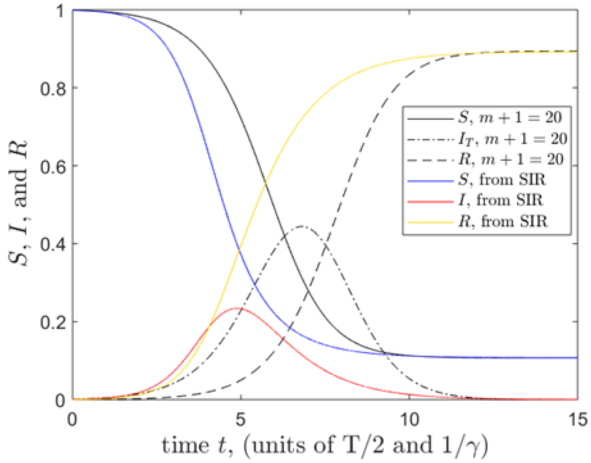

Figure 3: Comparing predictions of the standard SIR model (colored lines) with the SIR model including time since infection (black lines) for an epidemic with the same value of $R_0 = 2.5$. Some differences in the two sets of curves can be partially explained by different rescalings of the time domain, but many differences (including the peak number of infected persons) are attributed to details of the various stages of infection encoded in $\beta(s)$.

and a delay in the rise of a population that can be considered recovered.

For diseases that are less infectious (i.e. lower values of R_0) it may be that the details of a time-since-infection model are superfluous. For example, for $R_0 = 1.3$ Figure 4 makes the same comparison as in Figure 3 and finds much better agreement between the two models. The main disparity is that the time-since-infection model has a larger number of infected persons at every time. However, this disparity vanishes if one instead compares the infectiousness of the infected populations, i.e.$\int_{-1}^{1} ds\beta(s)I(t,s)$. Indeed, when the epidemic time is much much longer than the infectious period, $\epsilon = R_0 - 1 \ll 1$, the Legendre polynomial expansion can be truncated at $m+1 = 2$. In this limit, one can show show that the standard SIR model maps to a model with time since infection to leading order.

Finally, we consider how the number density of infected persons within the infectious range varies with time since infection when $R_0 = 2.5$. Early on, most infected persons are in the early stages of infection. This is confirmed in Figure 5, where we see that at $t = 1$ the number of new cases exceeds the number of expiring cases by a factor of seven:

Near the peak of the epidemic, the number of susceptibles has fallen into decline, limiting the number of new infections that can develop. At this time, the infectious population is somewhat uniformly distributed across the infectious window, as clearly shown in Figure 6.

In the late stages of the epidemic, most of the infected people are nearly

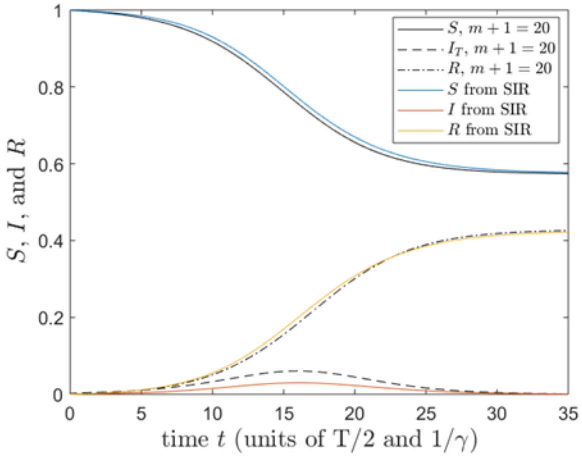

Figure 4: Comparing predictions of the standard SIR model (colored lines) with the SIR model including time since infection (black lines) for an epidemic with the same value of $R_0 = 1.3$. Overall, there is excellent agreement between the two model predictions. A key distinction is that the standard SIR model seems to predict a lower number of infectious at all times due to the non-constant value of $\beta(s)$.

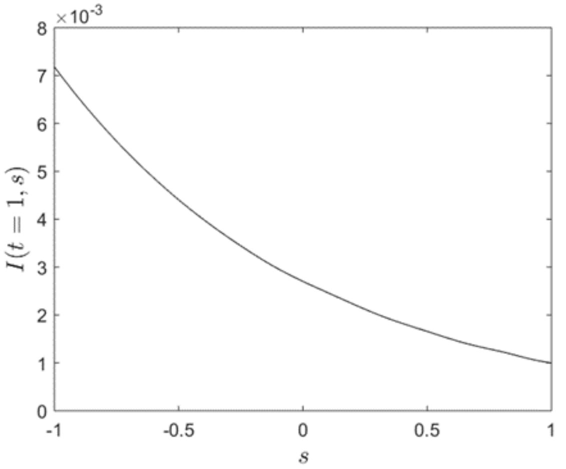

Figure 5: Predictions for the number density of infected persons for times since infection within the infectious range, $s \in [-1, 1]$ during an early stage of the epidemic, $t = 1$. In this early stage of the epidemic, the infectious population is primarily made up of relatively new cases.

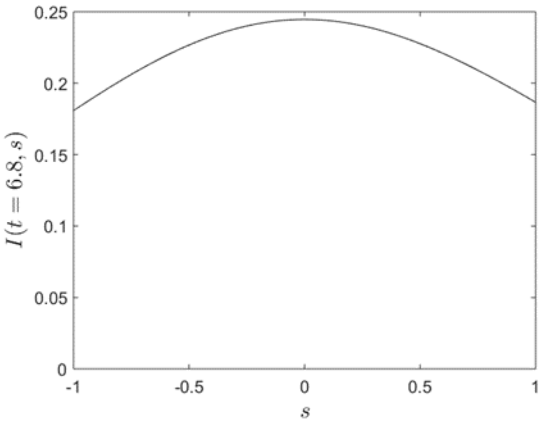

Figure 6: Predictions for the number density of infected persons for times since infection within the infectious range, $s \in [-1, 1]$ at the peak of the epidemic, $t = 6.8$. In this stage of the epidemic, the infectious population is almost uniformly spread across all stages of the disease.

recovered and there are comparatively few are new cases, as shown in Figure 7. This is the opposite of what was seen during the early stages of the epidemic.

Considering the curves shown in Figure 4 - Figure 6, it is clear that a low order polynomial fit is sufficient to describe the number density of infected persons in the infectious range, $s \in [-1, 1]$. This should generally be true as long as the epidemic progresses through the population on timescales that are much longer than the time for which an individual remains infectious. Since the timescale of the epidemic depends on R_0, higher values of R_0 require higher order polynomial interpolations to obtain good results.

At any time t in the calculation, one is always tracking a closed window of $I(t, s)$ in the range$s \in [-1, 1]$. The whole curve for a number density of persons infected at times $s > 1$ in the past can be reconstructed by splicing those windows together. For example, the simulation in Figure 2 ends at $t = 20$, at which point the entire number density curve $I(20, s)$ is reconstructed in Figure 8. Note that the peak around $s = 13$, corresponding to infections that developed at time $t = 20 - (s + 1) = 6$, is consistent with the peak in the number of infectious persons around $t = 7$ seen in 2.

The results of this section can be replicated using any of the aforementioned discretization strategies. As an example, we reproduce Figure 8 using the predictor/correcter method with a relatively sparse $N = 5$ stages with the correspondingly long time-step $h = 1/2$. In Figure 9 predictions from the Galerkin method are shown as the solid black line, while predictions from the predictor/correcter method are shown as open circles.

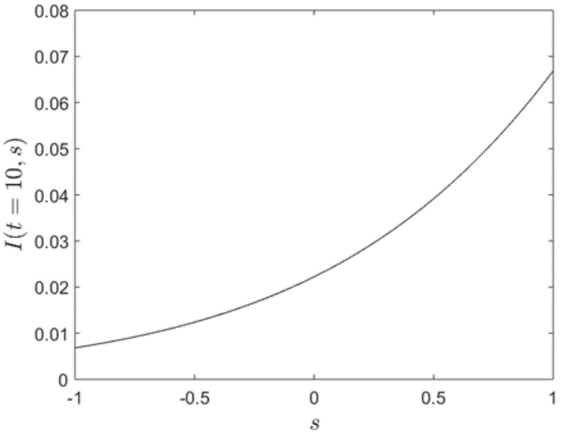

Figure 7: Predictions for the number density of infected persons for times since infection within the infectious range, $s \in [-1, 1]$ during a late stage of the epidemic, $t = 10$. In this stage of the epidemic, the infectious population is primarily made up of expiring cases.

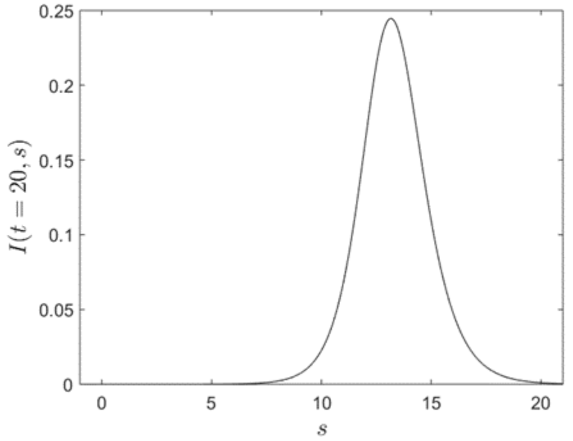

Figure 8: Predictions for the number density of persons infected at a time $s+1$ in the past, evaluated at $t = 20$. Note that the peak around $s = 13$, corresponding to infections that developed at time $t = 20 - (s + 1) = 6$, is consistent with the peak in the number of infectious persons around $t = 7$ seen in Figure 2.

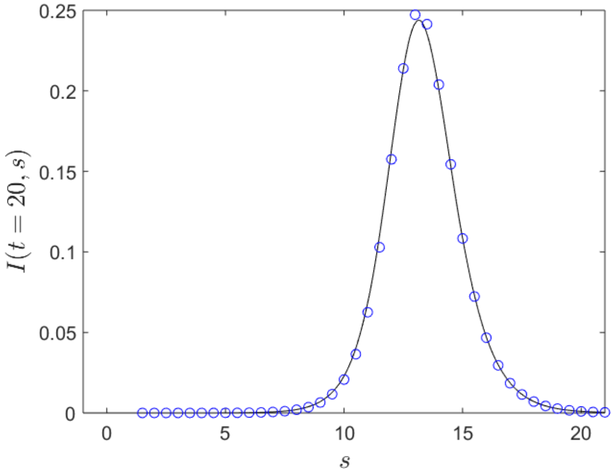

Figure 9: Comparing predictions from the Galerkin discertization (solid curve, c.f. Figure 8) with predictions from the predictor correcter method (blue circles).

3.3 Effect of a time-varying reproduction number

When the reproduction number R_0 can be treated as a constant, we have shown that converged predictions for $I(s,t)$ can be obtained with only a small number of Legendre polynomials. However, when the reproduction number R_0 varies in time, fluctuations in the reproduction number will create peaks and valleys in $I(s,t)$ corresponding to periods of high/low reproduction, respectively. In a practical sense, time-dependent changes in $R_0(t)$ could reflect patterns in human interaction due to cycles of night/day, weekday/weekend, and winter/summer as well as changes in social distancing policies, for example. Resolving short-time fluctuations in $I(s,t)$ requires both (1) a higher number of Legendre polynomials and (2) an implicit (as opposed to explicit) treatment of the boundary condition constraint Figure 41 whenever the Galerkin discretization is employed. It is the latter of these that poses the greater potential concern if the available ODE solvers are fully explicit.

When fluctuations in $R_0(t)$ occur on timescales that are very very short compared to T, one can use a moving average of the reproduction number to smooth out its fluctuations. This is generally an acceptable approximation as long as the smoothing time is small compared to the time-scale over which changes develop in $\beta(s)$. It is therefore fluctuations on timescales comparable to or longer than T (such as the cycle of weekday/weekend) that are most important to consider. To that end, we will present some sample calculations with time-varying R_0 of the form $R_0(t) = 2.5 + \Delta R_0 \sin(\omega t)$ and $\beta(s) \sim (1+s)^2(1-s)^2$. Once again, these results are presented as proof-of-concept for the SIcR model and do not purport to describe the progression of any particular disease.

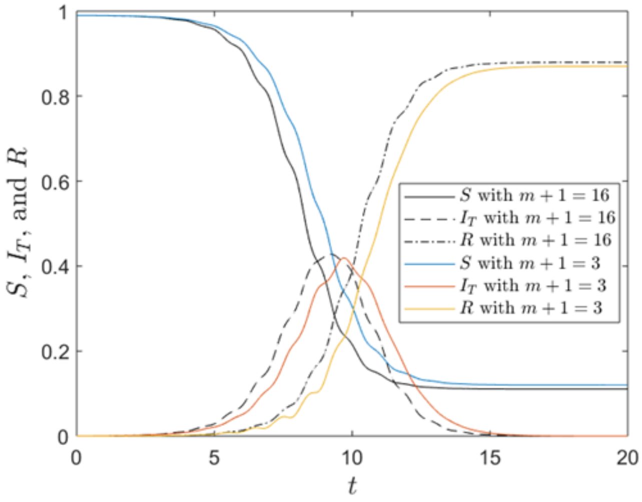

Figure 10: Comparing predictions with $m+1=16$ (converged) and $m+1=3$ for the case of a time-varying $R_0 = 2.5(1+0.5\sin(2\pi t))$. The small-mode approximation does capture the broad qualitative features of the converged results, but quantitative accuracy is poor. Around $t=7$, the small mode approximation even predicts a decreasing number of recovered cases.

First, we consider a fixed frequency and amplitude of oscillations, $\omega = 2\pi$ and $\Delta R_0 = 1.25$ to show that the presence of oscillations at short times increases the number of Legendre polynomials needed for converged results. In this case, we treat equation 41 implicitly and solve numerically in Matlab using ode15s. Figure 10 shows predictions for a small-mode approximation, m+1=3 and a fully converged calculation, m+1=16. Overall, the small-mode approximation does capture the broad qualitative trends of the converged results, but quantitative accuracy is poor in comparison to what is typically seen without oscillations (c.f. Figure 2).

When oscillations are present in $R_0(t)$ on time-scales less than T, higher order Legendre polynomials are required for tracking and preserving oscillations in $I(t,s)$. For the two simulations considered in Figure 10, we can perform the same analysis done for Figure 7 in an attempt to reconstruct the full I(t,s) at $t=20$ when the simulation is complete. In Figure 11, we sample the predicted $I(t, s \in [-1,1])$ at uniform increments of $t = 0, 1, 2, \ldots, 20$ and display all curves concurrently, shifted to $t=20$. In each case, the quality of convergence does not depend on the time interval at which the solution is sampled. The calculation can be considered converged if the curves spline together smoothly, otherwise the solution does not contain a sufficient number of Legendre polynomials.

Next, we show that when the boundary condition constraint of equation 41 is not handled implicitly (as it was in the preceding calculations), the same number of Legendre polynomials $m+1=16$ gives a worse approximation of

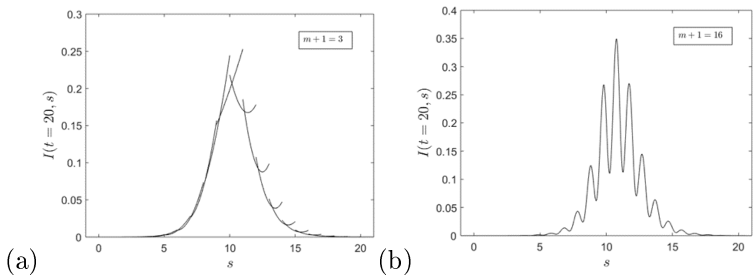

Figure 11: Comparing predictions for $I(t = 20, s)$ with a varying number of Legendre polynomials, $m+1 = 3$ (a) and $m+1 = 16$ (b) when the reproduction number varies in time as $R_0(t) = 2.5(1 + 0.5\sin(2\pi t))$. Note that oscillations in $R_0(t)$ create commensurate oscillations in $I(t, s)$ that require higher order Legendre polynomials to resolve. Using $m + 1 = 3$ Legendre polynomials is not sufficient for a numerically converged solution.

$I(t, s)$ depending on the times at which $I(s, t)$ is sampled. For this comparison, we solve the same problem numerically in Matlab, now using ode45. As shown in Figure 12, if we sample the solution at regular intervals $t = 0, 1, 2, \ldots, 20$ (black lines) we get a smooth curve. More generally, however, if we superimpose information sampled at time $t = 10.2$ we see a significant departure from the smooth curve at the end-points of the polynomial approximation (blue line). We will refer to the regions where these curves agree as a 'backbone' solution and the end-regions where they differ as 'fringe' solutions.

Fortunately, the 'fringes' that form along the curve of $I(s, t)$ do not appear to prevent us from observing the essential 'backbone' of the full curve for $I(s, t)$. Moreover, these 'fringes' (being limited to a narrow range near $s = \pm 1$) do not strongly contribute to the overall infectiveness of the infected class. Therefore, we can still resolve the progression of S, I_T, and R through the course of the epidemic. When using $m + 1 = 16$ modes, the results obtained using ode45 are identical (within the line thickness) to what is shown for the converged solution in Figure 10.

As a matter of best practice, a comparison of Figure 11(b) and Figure 12 shows that when the reproduction number varies in time, the boundary condition constraint from equation 41 should be handled implicitly to eliminate 'fringes' in the predictions for $I(t, s)$. In many cases (as in the examples shown here) these 'fringes' have no real impact on the model's overall predictions, but once the SIcR model is embedded into a model framework that includes geographic and demographic information it is difficult to say, a priori, what the impact of the 'fringes' will be.

Finally, we consider the effect of averaging out the time-dependence of the reproduction number altogether. In the curves below, we compare predictions for S, I_T and R given a time-dependent reproduction number $R_0(t) = 2.5 + 1.25\sin(2\pi t)$ and a constant reproduction number with the same mean value,

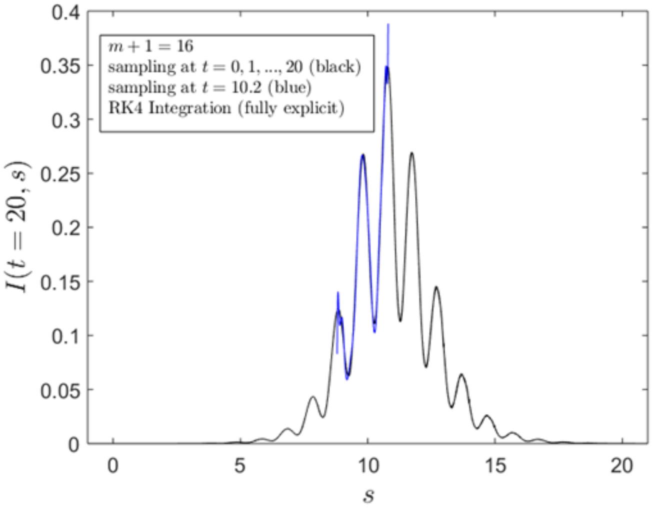

Figure 12: Predictions from a fully explicit solver for $I(t = 20, s)$ with $m+1 = 16$ when the reproduction number varies in time as $R_0(t) = 2.5(1 + 0.5\sin(2\pi t))$ constructed by sampling the solutions for $I(t, s)$ obtained at $t = 0, 1, 2, \ldots, 20$. When sampled at non-uniform times, however, there is less agreement.

$R_0 = 2.5$. From the comparison in Figure 13, we see that large fluctuations in R_0 occuring on time-scales comparable to but smaller than T only translate to small oscillations in the solutions for $S, I_T,$ and R. Therefore, we conclude that time-dependent changes in the reproduction number R_0 are only important to model to the extent that they change the mean value of R_0 on time-scales longer than T (as would occur during a lock-down, for example).

3.4 Hybrid Methods: Simulation of an abrupt Lockdown

To simulate an abrupt lockdown, we define a time-dependent$R_0(t)$ that drops abruptly at the time t_L when a lockdown is imposed. Here, we will present an example case in which $R_0(t)$ =2.5 for $t \leq t_L$and $R_0(t) = 0.7$ for $t > t_L$, with a lockdown imposed at t_L =3 following an initial condition of $I(0, s) = 10^{-3}\exp(-\lambda s)$ corresponding to the fastest growing linear mode for $R_0 = 2.5$ and $\beta(s) \sim (1-s)^2(1+s)^2$. A Galerkin discretization with $m + 1 = 10$ modes will be used for time $t \in [0, t_L]$ and $t \in [t_L + 2, 20]$ and the method of lines discretization with $N = 30$ will be used for the first infectious period following the abrupt change in R_0, $t \in [t_L, t_L + 2]$. In Figure 14, we plot the results for $I(t = 20, s)$ at the end of the simulation. Results from the Galerkin method are shown in black, while results form the method-of-lines discretization are shown in blue. Once again, these results are presented as proof-of-concept only and do not purport to describe the progression of any particular disease.

As a brief comment on Figure 14, it may initially come as a surprise to see

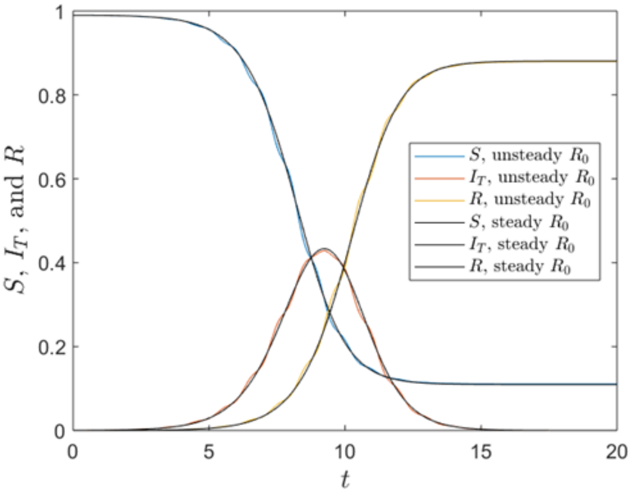

Figure 13: Comparing converged predictions for the SIcR model with constant reproduction number $R_0 = 2.5$ (solid lines) with a time-varying reproduction number $R_0 = 2.5(1 + .5\sin(2\pi t))$. The large oscillations in R_0 only translate to a small oscillation about the solution obtained with a constant reproduction number of the same mean value $R_0 = 2.5$.

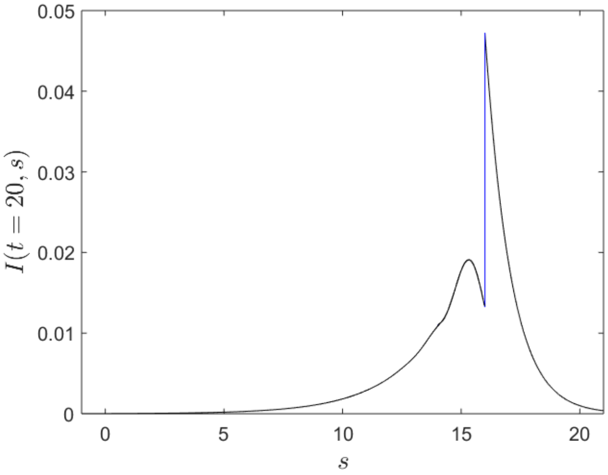

Figure 14: Predictions for the number density of infecteds in a simulation where there is an abrupt change in the reproduction number (as might occur with an abrupt lockdown). The signature of a lockdown is visable as a sharp discontinuity in $I(t, s)$ centered at $s = 16$ for $t = 20$, or more generally, at $t - (s + 1) = t_L$.

a continued increase in the number of infections even after the reproduction number has dropped below one. However this happens because prior to the lockdown most infections were too recent to be highly infectious. As those infections age and become more infectious, the number of new infections also rises. Once the pre-lockdown infections age beyond the time where they are most infectious, the number of new cases steadily drops again.

4 Discussion and Conclusions

Epidemic models are useful tools in the fight against infectious diseases, but they are often forced to strike a balance between accuracy and complexity. In the present work, we show that the accuracy of the SIR model can be improved at very low cost by accounting for infectiousness as a function of time since infection. We developed the governing PDE model, nondimensionalized, and reframed the equations (via a Galerkin approximation) into an ODE model whose structure closely aligns with existing generalizations of the SIR model [5, 1, 3, 8].

For R_0 close to unity, the predictions of the model with time-since-infection were qualitatively aligned with those of the standard SIR model. However, for more infectious diseases (i.e. larger R_0) where epidemic modelling is especially important, significant differences in the two frameworks were more evident. Therefore, it is our view that the SIR model with time-since-infection (as presented here) should be preferred to the standard SIR model for its more realistic description of disease transmission at virtually no added computational cost.

References

[1] L. Sattenspiel and K. Dietz, "A structured epidemic model incorporating geographic mobility among regions," Mathematical biosciences, vol. 128, no. 1, pp. 71 - 92, 1995.

[2] F. Hoppensteadt, "An age dependent epidemic model," Journal of the Franklin Institute, vol. 297, no. 5, pp. 325 - 333, 1974.

[3] D. Balcan, B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza and A. Vespignani, "Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model," Journal of computational science, vol. 1, no. 3, pp. 132 - 145, 2010.

[4] S. Busenberg, N. Mimmo Iannelli and H. R. Thieme, "Global behavior of an age-structured epidemic model," SIAM Journal on Mathematical Analysis, vol. 22, no. 4, pp. 1065 - 1080, 1991.

[5] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," Proceedings of the royal society of london. Series

A, Containing papers of a mathematical and physical character, vol. 115, no. 772, pp. 700 - 721, 1927.

[6] Diekmann, Odo and R. Montijn, "Prelude to Hopf bifurcation in an epidemic model: analysis of a characteristic equation associated with a nonlinear Volterra integral equation," Journal of Mathematical Biology, vol. 14, no. 1, pp. 117 - 127, 1982.

[7] A. Grant, "Dynamics of COVID-19 epidemics: SEIR models underestimate peak infection rates and overestimate epidemic duration," medRxiv, 2020.

[8] O. Krylova and D. Earn, "Effects of the infectious period distribution on predicted transitions in childhood disease dynamics," Journal of The Royal Society Interface, vol. 10, no. 84, 2013.

[9] A. Murillo and J. Alturo, "An epidemic model structured by the time since last infection," thesis, Purdue University, 2013.

[10] A. Lloyd, "Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics," Theoretical population biology, vol. 60, no. 1, pp. 59 - 71, 2001.

A Class-structured SIcR model

A.1 Dimensional Governing Equations

In many cases, it is important to resolve predictions for how an epidemic will spread through different sub-populations (e.g. age, occupation, socioeconomic status, etc.). If there are C classes in the population, we can denote population that is susceptible and assigned to class $i = 1, 2, \ldots, M$ as S_i. Likewise, in this appendix the population that's assigned to class i and was infected during the narrow interval of$s \in [s, s+\delta s]$ in the past is $I_i(t,s)\delta s$, and the total population assigned to class i is N_i. The mean frequency of contact between members of class i and j is given by the quantity C_{ij}/N_j, where $C_{ij}$is referred to as the 'contact matrix'. In the most general case, we allow each class i to have a different value of $\beta_i(s)$. The full PDE version of the governing equations are given by:

$$\frac{dS_i}{dt} = -S_i \sum_{j=1}^{M} \frac{C_{ij}}{N_j} \int_0^T ds \beta_j(s) I_j(t,s) \tag{46}$$

$$\frac{\partial}{\partial t} I_i(t,s) + \frac{\partial}{\partial s} I_i(t,s) = 0 \tag{47}$$

$$I_i(t,0) = S_i \sum_{j=1}^{M} \frac{C_{ij}}{N_j} \int_0^T ds \beta_j(s) I_j(t,s) \tag{48}$$

Sub-classes of the infected population (e.g. infectious, recovered, hosptialized, symptomatic, asymptomatic, etc.) can be computed in the same manner as before. For example, if we assume that all infecteds are classified as recovered once their infection ages beyond time T, then the fraction of the population that is recovered and assigned to sub-class i evolves as:

$$\frac{dR_i}{dt} = I_i(t, T) \tag{49}$$

A.2 Dimensionless Governing Equations

As before, we prefer to work with the governing equations in a dimensionless form. We rescale S_i and R_i by the total population $N = \sum_i N_i$, such that $\tilde{S}_i = S_i/N$ and $\tilde{R}_i = R_i/N$. Using a characteristic time $T_C = T/2$, we rescale time as $\tilde{t} = t/T_C$ and number density of infecteds as $\tilde{I}_i = I_i C/N$. The time since infection is rescaled to the variable $\tilde{s} = s/T_C - 1$ so that the range $s \in [0, T]$ is mapped to $\tilde{s} \in [-1, 1]$. Finally, we define a class-specific reproduction number

$$R_{0,i} = \sum_{j=1}^{M} \frac{C_{ij}}{N_j} \int_0^T ds \beta_j(s) \tag{50}$$

and a dimensionless matrix of rate constants that subsumes information about the contact structure:

$$\tilde{\beta}_{ij}(s) = \frac{T_C}{R_{0,i}} \frac{C_{ij}}{N_j} \beta_j(s) \tag{51}$$

Note that the matrix $\tilde{\beta}_{ij}$ has the property:

$$\sum_{j=1}^{M} \int_{-1}^{1} d\tilde{s} \tilde{\beta}_{ij}(\tilde{s}) = 1 \tag{52}$$

For compactness of notation, we will suppress tildes in our dimensionless variables – from this point forward, all variables are dimensionless unless otherwise stated. The dimensionless governing equations for the age-structured TSI model are given by:

$$\frac{dS_i}{dt} = -R_{0,i} S_i \sum_{j=1}^{M} \int_{-1}^{1} ds \beta_{ij}(s) I_j(t, s) \tag{53}$$

$$\frac{\partial}{\partial t} I_i(t, s) + \frac{\partial}{\partial s} I_i(t, s) = 0 \tag{54}$$

$$I_i(t, -1) = R_{0,i} S_i \sum_{j=1}^{M} \int_{-1}^{1} ds \beta_{ij}(s) I_j(t, s) \tag{55}$$

$$\frac{dR_i}{dt} = I_i(t, 1) \tag{56}$$

As before, there are many valid discretizations of the PDE model but as before we will focus on the Galerkin approximation and the method of lines.

A.3 Discretization: Method of Lines

Once again, we discretize s in uniform intervals of width h and represent advection/integration in s via upwinding and left Riemann sums, respectively. We evaluate the solution at discrete points in s, evaluating at $s_1, s_2, s_3 \ldots, s_N$ where $s_n = -1 + (n-1)h$ and $h = 2/(N-1)$. In this section, we denote $I_{i,n}$ as the number density of infected persons at each s_n and $\beta_{ij,n} = h\beta_{ij}(s_n)$. Applying this discretization scheme to the PDE equations, we obtain evolution equations for $S_i, I_{i,n}$, and R_i:

$$\frac{dS_i}{dt} = -R_0 S_i \sum_{n=2}^{N} \beta_{ij,n} I_{i,n} \tag{57}$$

$$I_{i,0} = R_0 S_i \sum_{n=2}^{N} \beta_{ij,n} I_{i,n} \tag{58}$$

$$\frac{dI_{i,n}}{dt} = \frac{1}{h}(I_{i,n-1} - I_{i,n}) \tag{59}$$

$$\frac{dR_i}{dt} = I_{i,N} \tag{60}$$

This maps exactly to the standard age-structured multi-stage SIkR model. Therefore, we confirm that in the limit of $h \to 0$, the age-structured SIkR model is equivalent to an age-structured time since infection model. In practice, however, the convergence is so slow that truncation errors will have a large influence on the predictions. As before, the truncation error can be represented as numerical diffusion in the $s-$direction.

As before, one can also employ the more efficient predictor/corrector discretization using the method of lines.

A.4 Discretization: Galerkin Approximation

We can represent each $I_i(t, s)$ on $s \in [-1, 1]$ via a truncated Legendre polynomial expansion with $m + 1$ polynomials:

$$I_i(t, s) = \sum_{n=0}^{m} I_{i,n}(t) P_n(s) \tag{61}$$

Inserting this into the PDE equations, we obtain:

$$\frac{dS_i}{dt} = -S_i \sum_{j=1}^{M} \sum_{n=0}^{m} a_{ij,n} I_{j,n} \tag{62}$$

$$a_{ij,n} = \int_{-1}^{1} ds \beta_{ij}(s) P_n(s) \tag{63}$$

$$\left[\frac{2}{2n+1} \right] \frac{d}{dt} I_{i,n} + \sum_{k=0}^{m} b_{nk,i} I_{i,n} = 0 \tag{64}$$

The coefficients $b_{nk,i}$ evaluate to $b_{nk} = 2$ if $n+k$ is odd and $k > n$, otherwise they evaluate to zero. The highest order Legendre Polynomial coefficient $I_{i,m}$is constrained to satisfy:

$$\sum_{n=0}^{m} I_{i,n}(-1)^n = S_i \sum_{j=1}^{M} \sum_{n=0}^{m} a_{ij,n} I_{j,n} \tag{65}$$

A.5 Linear Stability Analysis

As a generalization of the linear stability analysis for the single compartment case, we suppose an initial condition in which nearly the whole population is susceptible and there is a small population of infected growing as $I_i(t, s) = I_{0,i} \exp(\lambda(t - s))$. The growth rate λ and relative number of infected in each age compartment $I_{0,i}$ can be obtained by considering the following matrix:

$$A_{ij} = R_{0,i} S_i \sum_{j=1}^{M} \int_{-1}^{1} ds \beta_{ij}(s) e^{-\lambda(s+1)} \tag{66}$$

The growth rate λ is the value of lambda for which the largest eigenvalue of A_{ij} is equal to one, and the relative number of infected in each age compartment $I_{0,i}$ is given by the corresponding eigenvector.