RESEARCH ARTICLE

Check for updates

On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods

Maziar Raissi[a], Niloofar Ramezani[b] and Padmanabhan Seshaiyer[c]

[a]Division of Applied Mathematics, Brown University, Providence, RI, USA; [b]Department of Statistics, George Mason University, Fairfax, VA, USA; [c]Department of Mathematical Sciences, George Mason University, Fairfax, VA, USA

ABSTRACT

In this paper, we consider compartmental disease transmission models and discuss the importance of determining model parameters that provide an insight into disease transmission and prevalence. After a brief review and comparison of the performance of some heuristic approaches, the paper introduces three approaches including an optimization approach, a physics informed deep learning and a statistical inference method to estimate parameters and analyse disease transmission. The deep learning framework utilizes the hidden physics of infectious diseases and infer the latent quantities of interest in the model by approximating them using deep neural networks. The performance of the deep learning method is validated against representative small and big data sets corresponding to a well-known benchmark example and the results indicate that deep learning is a viable candidate to determine model parameters. The paper also presents the need for researchers to understand different types of dependencies exhibited in a typical data set and discovering the most appropriate approaches for statistical inference. Specifically, in this work we apply a time-series inferential method with a variety of statistical models. Our results indicate the efficiency and importance of statistical inference methods for researchers to understand and analyse time-series data to make confident predictions.

1. Introduction

Understanding the dynamics of infectious diseases using mathematical models can be traced back to first epidemic models proposed by Kermack and McKendric in 1927 (Brauer & Castillo-Chavez, 2012; Kermack & McKendrick, 1927). This original compartmental model assumed that the total population N may be divided into three distinct classes of sub-populations S, I and R denoting the *susceptible*, *infected* and *recovered* classes respectively. The *susceptible* class of individuals included members of the population that have the potential to contract a disease and their size is denoted by S. The *infected* class of individuals are those that are assumed to have contracted the disease and this class is denoted
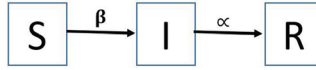

Figure 1. Compartmental model for SIR model.

by I. The final class of individuals denoted by R consisted of those that *recovered* and cannot contract the disease again. Further, it is also assumed that the number of individuals in each of these classes (*compartments*) change with time, that is, $S(t)$, $I(t)$ and $R(t)$ are functions of time t and the total population N is the sum of the number of individuals in these compartments. Hence,

$$N = S(t) + I(t) + R(t). \tag{1}$$

Since models represent approximations to reality, one must make suitable assumptions to simplify reality. In the development of the mathematical model, two principal assumptions were made in the Kermack McKendrick model that included that infected individuals are also infectious and that the total population size N remains constant. This simple SIR epidemic model can be illustrated in compartments as in Figure 1. Not only was it capable of generating realistic single-epidemic outbreaks but also provided important theoretical epidemiological insights.

In Figure 1, it is assumed that each class resides within exactly one compartment and can move from one compartment to another. Each compartment in the figure is represented by a box indexed by the name of the class and arrows denote the direction of movement of individuals between the classes. When a susceptible individual enters into contact with an infectious individual, then that susceptible individual becomes infected with a certain probability and moves from the susceptible to the infected class. Therefore, the susceptible population decreases in unit of time by all individuals who become infected in that time while at the same time the class of infectives will be assumed to increase by the same number of newly infected individuals. The dynamics of the three sub-populations $S(t)$, $I(t)$ and $R(t)$ may be described by the following SIR model given by first-order coupled ordinary differential equations (ODE):

$$\begin{aligned}
\frac{dS}{dt} &= -\beta \, S \, I, \\
\frac{dI}{dt} &= \beta \, S \, I - \alpha \, I, \\
\frac{dR}{dt} &= \alpha \, I.
\end{aligned} \tag{2}$$

Note that this closed system does not allow any births/deaths. This SIR model in system (2) is fully specified by prescribing the *transmission* and *recovery* rates along with a set of initial conditions $S(0)$, $I(0)$ and $R(0)$. The total population N at time $t = 0$ is given by $N = S(0) + I(0) + R(0)$. Adding all the equations in system (2), we notice that $N'(t) = 0$ and therefore $N(t)$ is a constant and equal to its initial value. One can further assume $R(0) = 0$ since no one has yet had a chance to recover or die. Thus a choice of $I(0) = I_0$ is enough to define the system at $t = 0$ since then $S_0 = N - I_0$. If the epidemic is triggered by a single infected individual, one can take $t = 0$ to be the moment at which $I = I_0 = 1$.

If we denote the parameter β, the *transmission rate*, then $\beta S I$ denotes the number of individuals who become infected per unit of time. It is also assumed that the individuals that recover or die leave the infected class at a constant per capita probability per unit of time which we denote as α, the *recovery rate*. This implies αI is the number of individuals per unit of time who recover and therefore leave the infectious class and move to the recovered class. It is very important to obtain precise estimates of these disease transmission and recovery rates for epidemiological simulation models. Most often these rates are estimated from longitudinal field data, which are costly and time-consuming to conduct (Backer, Berto, McCreary, & Martelli, 2012; Caley & Ramsey, 2001). For most models, these parameters are hard to estimate because natural processes are stochastic, and transmission events are influenced by other parameters than what can be included in a transmission model (Anderson & May, 1992; Elkadry, 2013). Often some of the parameters such as the *recovery rate* maybe estimated from patterns in the data but *transmission rates* have to be computed using heuristic algorithms that are computationally or statistically motivated. Some of these computational algorithms include inverse methods, least-squares approach, agent-based modelling, using final size calculations (Martcheva, 2015; Murray, 1989; Pollicott, Wang, & Weiss, 2012; Yong, Mubayi, & Kribs, 2015). Also, researchers have employed a variety of statistical approaches including maximum-likelihood, Bayesian inference and Poisson regression methods (Capaldi et al., 2012; Hadeler, 2011; Huang, Liu, & Wu, 2006; Longini, Koopman, Haber, & Cotsonis, 1988; O'Dea, Pepin, Lopman, & Wilke, 2014). Some of this work also showed that the precision of the estimate increased with the number of outbreaks used for estimation (O'Dea et al., 2014). To determine the relative importance of model parameters to disease transmission and prevalence, there has also been work around *sensitivity analysis* of the parameters using techniques such as Latin Hypercube Sampling and Partial Rank Correlation Coefficients analysis with the associated mathematical models (Blower & Dowlatabadi, 1994; Chitnis, Hyman, & Cushing, 2008; McKay, Beckman, & Conover, 1979).

Recently, there has been some new work around probabilistic machine learning to discover governing equations expressed by parametric linear operators (Raissi & Karniadakis, 2018; Raissi, Perdikaris, & Karniadakis, 2017a, 2017b, 2017c, 2017d). Such equations involve, but are not limited to, ordinary and partial differential, integro-differential and fractional order operators. A grand challenge in mathematical biology and epidemiology with great opportunities facing researchers working on infectious disease modelling is to develop a coherent deep learning framework that enables them to blend differential equations such as the system (2) with the vast data sets now available. One of the tools that makes these deep learning methods successful is the use of *neural networks* which is a system of decisions modelled after the human brain (LeCun, Bengio, & Hinton, 2015). Consider the illustration shown in Figure 2. The first layer of perceptrons first weigh and bias the input which can be observed values of infected data. The next layer then will make more complex decisions based off those inputs, until the final decision layer is reached which generates the outputs which can correspond to the values of parameters such as β and α. In this research, we implement a physics informed neural network-based approach which makes decisions based on appropriate activation functions depending on the computed *bias* (b) and *weights* (w). The network then seeks to minimize the mean squared error of the regression with respect to the weights and biases by utilizing gradient descent type methods used in conjunction with software such as tensorflow. While there is currently
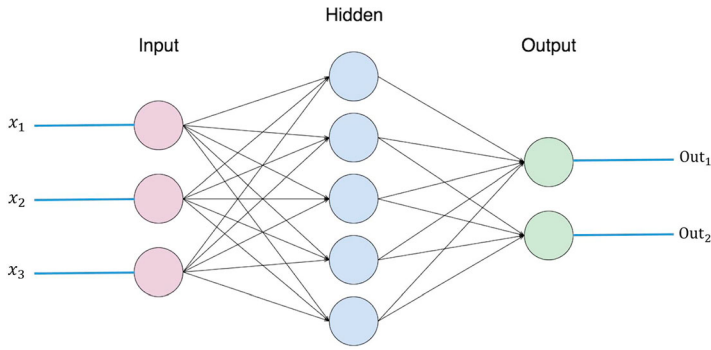
Figure 2. An illustration of a neural network.

a lot of enthusiasm about 'big data', useful data in infectious diseases is usually 'small' and expensive to acquire. In this work, we will describe how one can apply such physics informed neural network-based deep learning approaches and apply it to a real-world example to estimate optimal parameters, namely the transmission and recovery rates, in the SIR model. This is one of the novel contributions of this work.

Mathematical models for systems such as (2) coupled with statistical inference techniques have allowed researchers to compare infectious disease theory and data which in turn has helped to provide insight on transmission estimates, vaccine control strategies and predicting future trends. Despite its simplicity, calibrating the SIR model against time-series data has been a challenging problem over the years (Finkenstädt & Grenfell, 2000; Wallinga, Lávy-Bruhl, Gay, & Wachmann, 2001). One of the reasons is the need for researchers to understand the different types of dependencies that are exhibited by the data. A second contribution of this work includes discovering the most appropriate approach to statistical inference to get a better understanding and analysis of the time-series data. Within this efficient statistical approach, which, to the best of our knowledge, has not been used before to investigate the trend of epidemic and infectious diseases, not only multiple sources of dependencies among individuals are taken into consideration to minimize estimation error and avoid making unreliable conclusions, but also the possibility of including transmission and recovery rates estimated from the real data is provided to build more accurate models. Such predictive models can accurately estimate the current trend of infection and predict the future trend, within the population of study. After building the aforementioned statistical model and evaluating it, through cross validation, it can be used to estimate the transmission and recovery rates, as well as infection trend, in other populations. This can greatly assist epidemiologists and other scientists who want to predict the infection rate of a disease in different regions and locations and be prepared for the occurrence and outbreak of it in newly exposed regions.

In this paper, we start with analysing the system (2) to determine formulas for final sizes for the susceptible and recovered populations. We then introduce one of the most commonly used benchmark examples with real-world data and discuss how these formulas can be used along with estimated values of transmission and recovery rates to predict and visualize the observed data. We illustrate how the work of a variety of researchers (Martcheva, 2015; Murray, 1989; Yong et al., 2015) applied to the same real-world data

end up predicting different transmission and recovery rates. Following this we also illustrate how one can predict these rates using an optimization algorithm that yields a better comparison between the real-world data and the computed data. In Section 3, we present a new paradigm of learning differential equations from small data for infectious diseases. In particular, we introduce hidden physics models, which are essentially data-efficient learning machines capable of leveraging the underlying laws of physics, expressed by time-dependent differential equations, to extract patterns from high-dimensional data generated from experiments. The proposed methodology may be applied to the problem of learning, system identification or data-driven discovery of ordinary differential equations such as the system (2). Our framework relies on Gaussian processes, a powerful tool for probabilistic inference over functions, that enables us to strike a balance between model complexity and data fitting. We note that our implementation of this new algorithm ends up predicting the transmission and recovery rates with minimal information. We believe this is the first time such a deep learning algorithm is applied in the context of infectious diseases. Section 4 provides some important tools and observations that researchers in this area should consider employing to analyse time-series data through the real-world example. Once again, we believe this is the first time such a time-series inferential method is applied in the infectious diseases context. Finally, in Section 5, we summarize the discussions in the work and conclude.

2. Estimating transmission and recovery rates

In this section, we introduce the concept of *final size* and derive explicit formulas to compute them for system (2). Following that we introduce the data from a popular real-world example of an influenza outbreak in a boys boarding school and introduce various techniques to estimate the transmission and recovery rates as suggested by various researchers. Finally, we present an optimization algorithm using Nelder–Mead that obtains optimal estimates for the transmission and recovery rates that let the computed values of infected data to be the closest to the observed values of the infected data.

2.1. Determining final size

Consider system (2). Note that from the first equation in system (2), $S'(t) < 0$ for all t and therefore $S(t)$ is monotonically decreasing. This implies that the number of susceptible individuals is always declining independently of the initial condition $S(0)$ and we also have

$$\lim_{t \to \infty} S(t) = S_\infty.$$

This quantity S_∞ is called the *final size* of the epidemic. One may also note that when $S = \alpha/\beta$ the second equation in system (2) to be zero, namely, $I'(t) = 0$. Hence, $I(t)$ can have a stationary point at some maximum time. On the other hand, the number of infected individuals may be monotonically decreasing to zero, or may have non-monotone behaviour by first increasing to some maximum level, and then decreasing to zero. One may note that the spread start to increase if $I'(0) > 0$ which yields the following necessary and

sufficient condition for an initial increase in the number of infectives given by

$$\frac{\beta\ S(0)}{\alpha} > 1.$$

Thus if $S_0 < \alpha/\beta$, the infection dies out and there is no epidemic.

The last equation in system (2) also indicates that the recovered individuals also have monotone behaviour, independent of $R(0)$. Since $R'(0) > 0$ for all t, the number of recovered is always increasing monotonically. Since we know that this number is constrained by the total population N, we also have

$$\lim_{t\to\infty} R(t) = R_\infty.$$

Dividing the first and third equations in system (2), one obtains

$$\frac{dS}{dR} = -\frac{\beta}{\alpha}\ S,$$

which can be solved to yield

$$S(t) = S(0)\ e^{-(\beta/\alpha)\ R}. \tag{3}$$

Since $R < N$ this gives

$$S(t) \geq S(0)\ e^{-(\beta/\alpha)\ N} > 0.$$

This implies that the final size $S_\infty > 0$.

Integrating the first equation of system (2), one can show that

$$S_0 - S_\infty \geq \beta\ S_\infty \int_0^1 I(t)\ dt.$$

This implies that $I(t)$ is integrable on $[0, \infty)$ and hence

$$\lim_{t\to\infty} I(t) = I_\infty = 0.$$

Since $N = S_\infty + R_\infty = S_0 + I_0$ is a constant, Equation (3) may be rewritten as

$$S_\infty = S_0 e^{-(\beta/\alpha)\ (S_0+I_0-S_\infty)}. \tag{4}$$

If the parameters α and β are known along with initial number of susceptible population $S(0)$, this equation maybe solved numerically. On the other hand, if we know the final size S_∞ and the total population N, one can use this equation to estimate β/α. Dividing the first and second equations of system (2), one can see that

$$\frac{dI}{dS} = -1 + \frac{\alpha}{\beta\ S}, \tag{5}$$

which yields

$$I + S - \frac{\alpha}{\beta} \ln S = K,$$

where K is an arbitrary constant. Assuming then that $I + S - (\alpha/\beta) \ln S = I_0 + S_0 - (\alpha/\beta) \ln S_0$, one can calculate the maximum number of infected individuals I_{max} reached

in the epidemic which occurs at $S = \alpha/\beta$:

$$I_{\max} = -\frac{\alpha}{\beta} + \frac{\alpha}{\beta} \ln\left(\frac{\alpha}{\beta}\right) + I_0 + S_0 - \frac{\alpha}{\beta} \ln S_0. \tag{6}$$

If we are able to estimate $I_{\max}$ for a newly occurring infectious disease, one can determine when the number of infections will begin to decline. Also, note that using Equation (5), one can compute

$$R_\infty = \frac{\alpha}{\beta} \ln\left(\frac{S_0}{S_\infty}\right). \tag{7}$$

2.2. Application to real data

In 1978, the British Medical Journal *Lancet* reported on 4 March 1978 about an outbreak of influenza virus in a boys boarding school (Anonymous, 1978). The school had a population of 763 boys, all of whom were at risk during the epidemic. One boy who had returned from an overseas trip is believed to have initiated an influenza epidemic in the school after his return. At the outbreak of the epidemic, none of the boys previously had influenza and so no resistance to infection was present. Of these 512 were confined to bed during the epidemic, which lasted from 22 January 1978 until 4 February 1978. This data is illustrated in Figure 3 that also plots the data of convalescent who recovered after the illness.

Note that several researchers have interpreted the exact values of the infected number of individuals extracted from the graph in Figure 3 very closely (Martcheva, 2015; Murray, 1989). We will use the representative dataset illustrated in Table 1 for our work. Based on what one assumes is known in the SIR model in system (2), one can predict a variety of useful quantities. One way to help realize the solution to the system (2) is to employ numerical
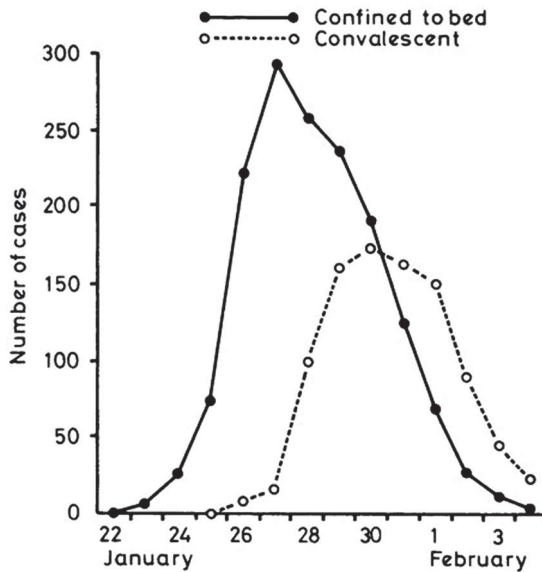

Figure 3. Sample influenza dataset epidemic in an English boarding school (Anonymous, 1978).

Table 1. Sample influenza dataset (Anonymous, 1978).

Day	Infected number
1	3
2	8
3	28
4	75
5	221
6	291
7	255
8	235
9	190
10	126
11	70
12	28
13	12
14	5

methods that help approximate the differential equations. Specifically, in this work given the estimated values of the parameters, we will employ the higher order Runge–Kutta methods to implement system (2) and study the dynamics of the three human sub-populations, namely, Susceptible, Infected and Recovered. Our first task in order to study the dynamics of system (2) is to estimate parameters β and α from the given data in Table 1.

2.2.1. Estimating parameters for unknown S_∞

One can use a simple (but approximate) approach to obtain the estimate for the parameters as follows. Since the epidemic was initiated by one sick boy infecting two more sick boys 1 day later, a crude approximation could be $S'(t) \approx -2$ per individual per day. With $S_0 = 762$ and $I_0 = 1$, one can estimate the initial transmission rate to be

$$\beta \approx \frac{-S'(t)}{S\,I} \approx \frac{2}{762 \times 1} = 0.0026.$$

The report indicated that the boys were taken to the infirmary within 1 or 2 days of becoming sick. So one may estimate that 1/2 of the infected population was removed each day or $\alpha = 0.5$ per day. This then yields the ratio of $\alpha/\beta = 192$. Using these values, one can then plot the dynamics of the model predictions compared to the data using the system (2). We employ a higher order Runge–Kutta method to implement system (2) and the dynamics of the three sub-populations are plotted in Figure 4. Moreover, one can use Equations (4) and (6) to determine $S_\infty = 16$ and $I_{max} = 306$.

Using the SIR model (2), Murray (1989) also reported performing a careful fit for the data in Table 1 to obtain $\alpha/\beta = 202$, $\beta = 0.00218$ per day. The initial conditions were assumed to be $N = 763$, $S_0 = 762$ and $I_0 = 1$. The dynamics of the model predictions compared to the data using the system (2) is plotted in Figure 5. Note also that, using (6) one can determine $I_{max} = 445$.

Note that the parameter values used to plot Figures 5 and 4 are close and predict a similar course for the disease. The conditions for an epidemic are clearly met in both cases according to the model since $S_0 > \alpha/\beta$. For our crude approximation, one can use Equation (7) to yield to $R_\infty = 747$.
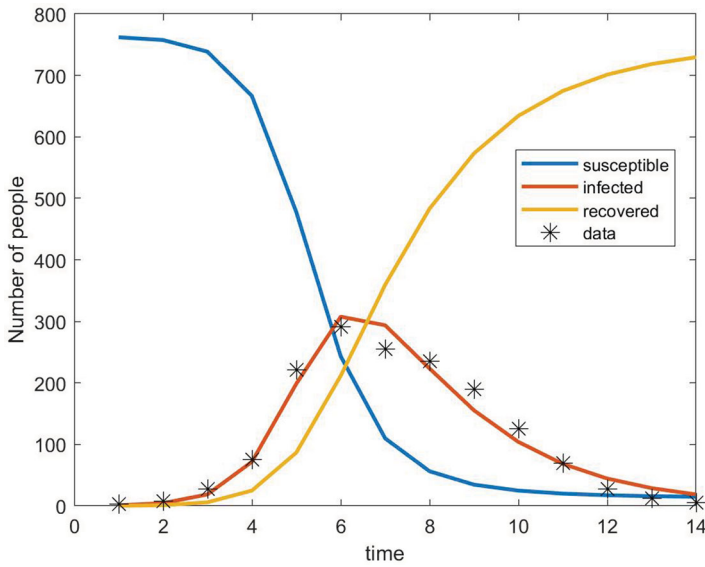
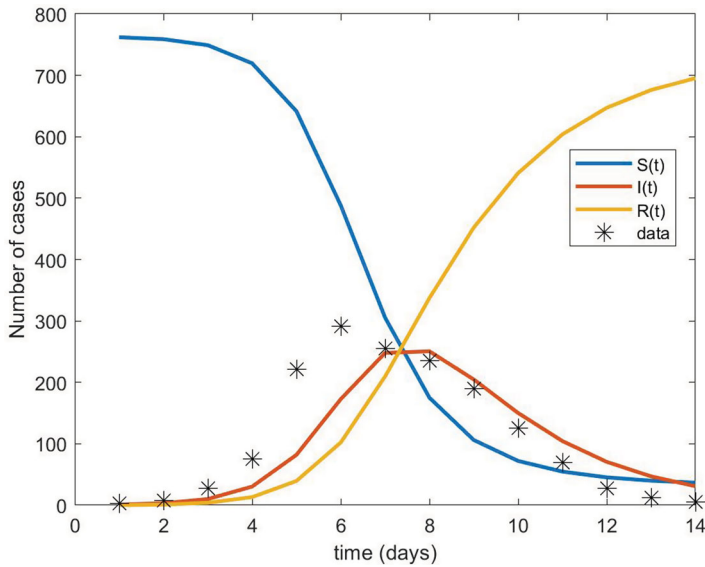**Figure 4.** Dynamics of SIR model predictions and comparison to data for $\alpha = 0.5$ and $\beta = 0.0026$.

Figure 5. Dynamics of the SIR model predictions and comparison to data for $\alpha = 0.5$ and $\beta = 0.00218$.

2.2.2. Estimating parameters for known S_∞

Another approach to determine the rates was employed by Martcheva (2015) who noted that number of boys who seems to have escaped the influenza epidemic was 19 which serves as S_∞. Using Equation (4) with the data, one can determine the ratio:

$$\frac{\beta}{\alpha} = \frac{\ln\left(\frac{762}{19}\right)}{763 - 19} = 0.00491704. \tag{8}$$
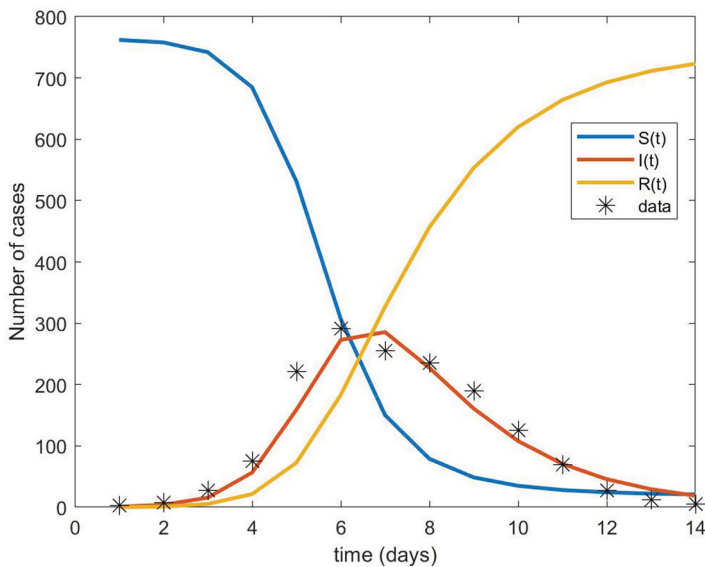
Figure 6. Dynamics of the SIR model predictions and comparison to data for $\alpha = 0.5$ and $\beta = 0.002481$.

Assuming as before that the boys were quarantined in the infirmary for about 2 days as infectious individuals, one can find the recovery rate to be $\alpha = \frac{1}{2} = 0.5$. From Equation (8), we can calculate the value of $\beta = 0.002481$. We can now use Equation (6) to determine $I_{\max} = 293.4$. Using these values, one can then plot the dynamics of the model predictions compared to the data using the system (2) which is plotted in Figure 6.

It must be pointed out that this data set consists of a closed population. It must be also noted that all these models assumed that the recovery rate α can be computed heuristically. However, there are also methods in the literature that can help to estimate the parameters including S_0, α and β in system (2) by minimizing the deviation between the SIR model out and a given data set. One such method is the Berkeley Madonna method (Macey, Oster, & Zahnley, 2000) which has been shown to fit the data using the fitting parameter as α which was estimated to be 0.4421 when $N = 763$ and $\beta = 0.00218$ (Yong et al., 2015). The dynamics corresponding to these parameters is illustrated in Figure 7.

Clearly, from Figures 4–7, there are variations in the ability of the dynamics of the computed values of infected population to track the true data for the various combination of parameters. As noted, these parameters were calculated through heuristic methods in some of these algorithms and may not be optimal. Next, we consider an optimization algorithm that employs a least-squares minimization approach to estimate optimal parameters.

2.3. Estimating optimal parameters α and β

One way to efficiently estimate the optimal parameters is to employ an unconstrained non-linear optimization algorithm such as the Nelder–Mead algorithm which searches for a local minimum using a regression approach. This direct search method attempts to minimize a function of real variables using only function evaluations without any derivatives
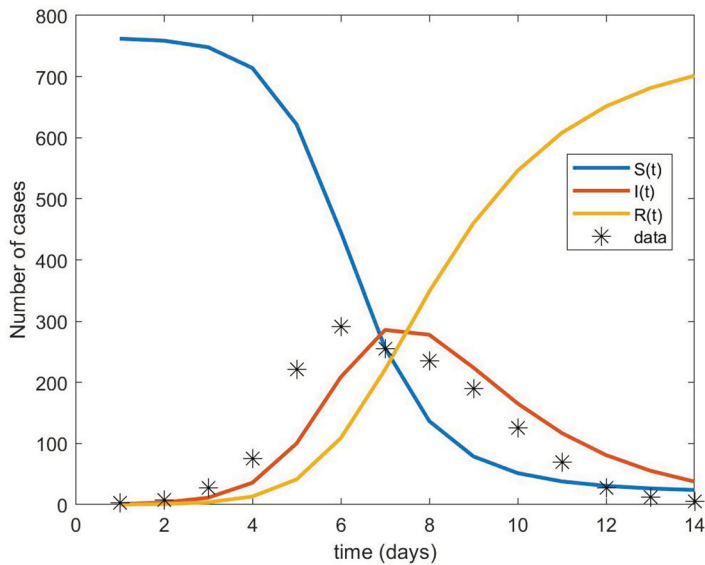
Figure 7. Dynamics of the SIR model predictions and comparison to data for $\alpha = 0.4421$ and $\beta = 0.00218$.

(Nsoesie, Beckman, Shashaani, Nagaraj, & Marathe, 2013). The minimized objective function is represented by differences in the daily infected counts from observed data and the computer simulated data. Using the initial conditions $S_0 = 762, I_0 = 1, R_0 = 0$ with a poor guess for $\alpha_{\text{guess}} = 0.1, \beta_{\text{guess}} = 0.01$, the optimization algorithm estimates the parameters to be $\beta = 0.002568$ and $\alpha = 0.4768$. The predicted dynamics of the three sub-populations for these parameter values along with comparison to the true data is illustrated in Figure 8.
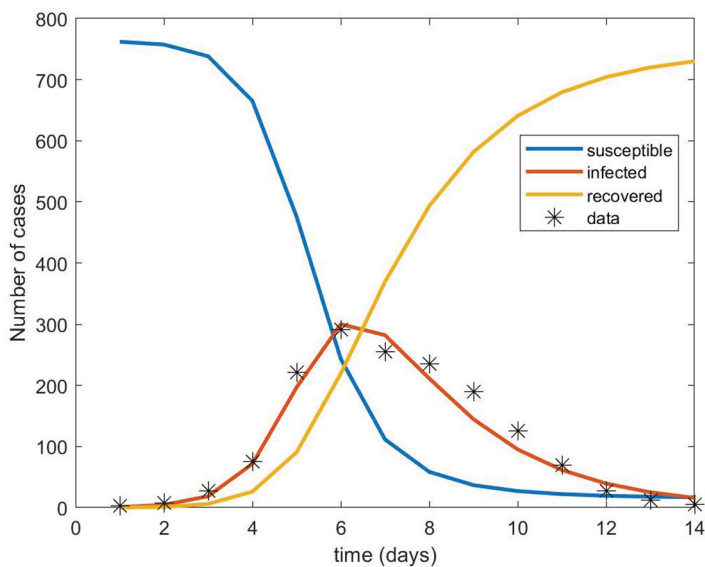

Figure 8. Dynamics of the SIR model predictions and comparison to data for $\alpha = 0.4768$ and $\beta = 0.002568$.

All the methods that have been discussed so far assumed the prior knowledge of initial number of each of the human sub-populations including the Susceptible S_0, Infected I_0 and Recovered R_0. Next, we introduce methods using physics informed deep learning that will not only be able to predict parameters α and β but also help predict initial susceptible population S_0 along with the dynamics for given infected data from Table 1 and $R_0 = 0$.

3. Physics informed deep learning approach

In the following, inspired by recent developments in *physics-informed deep learning* (Raissi et al., 2017c, 2017d) and *deep hidden physics models* (Raissi & Karniadakis, 2018; Raissi et al., 2017a, 2017b), we propose to leverage the hidden physics of infectious diseases (i.e. Equations 1 and 2) and infer the latent quantities of interest (i.e. S, I and R) by approximating them using deep neural networks. This choice is motivated by modern techniques for solving forward and inverse problems associated with differential equations, where the unknown solution is approximated either by a neural network (Raissi et al., 2017c, 2017d; Raissi, Perdikaris, & Karniadakis, 2018) or a Gaussian process (Raissi & Karniadakis, 2018; Raissi et al., 2017a, 2017b). Following these approaches, we approximate the latent function

$$t \longmapsto (S, I, R)$$

by a deep neural network and obtain the following *physics informed neural networks* (see Figure 9) corresponding to Equations (1) and (2), i.e.

$$
\begin{aligned}
E_1 &:= \frac{dS}{dt} + \beta\, S\, I, \\
E_2 &:= \frac{dI}{dt} - \beta\, S\, I + \alpha\, I, \\
E_3 &:= \frac{dR}{dt} - \alpha\, I, \\
E_4 &:= N - (S + I + R).
\end{aligned}
\tag{9}
$$

A schematic representation of the resulting *physics informed neural networks* is given in Figure 9. Note that for simplicity of illustration, Figure 9 depicts a network that comprises of two hidden layers and four neurons per hidden layers. Networks with this kind
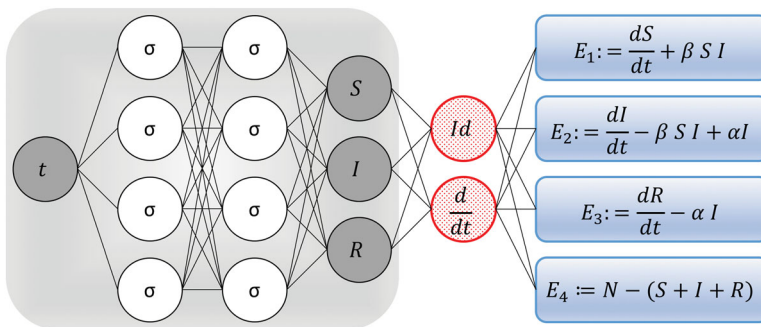

Figure 9. Physics informed neural networks.

of many-layer structure – two or more hidden layers – are called *deep neural networks*. These neurons in the network may be thought of as holding numbers that are calculated by a special *activation function* that depends on suitable *weights* and *biases* corresponding to each connection between neurons in each layer. With prior knowledge of such an activation function, the problems boil down to identifying the weights and biases that correspond to computed values of infected data that is close to the observed values. The three sub-populations are approximated by the deep neural network with calculus on computation graphs using a backpropagation algorithm (Goodfellow, Bengio, Courville, & Bengio, 2016; Hecht-Nielsen, 1992; Schmidhuber, 2015).

We acquire the required derivatives to compute the residual networks E_1, E_2, E_3 and E_4 by applying the chain rule for differentiating compositions of functions using automatic differentiation (Baydin, Pearlmutter, Radul, & Siskind, 2018). In our formal computations, we employed a densely connected (physics uninformed) neural network, with 1 hidden layer and 32 neurons per hidden layer which takes the input variable t and outputs S, I and R. The activation function we employed was

$$\sigma(x) = \tanh(x). \tag{10}$$

We employ automatic differentiation to obtain the required derivatives to compute the residual (physics informed) networks E_1, E_2, E_3 and E_4. It is worth highlighting that parameters α and β of the differential equations turn into parameters of the resulting physics informed neural networks $E_1 - E_4$. The total loss function is composed of the regression loss corresponding to I and the loss imposed by the differential equations system (9). Here, *Id* denotes the identity operator and the differential operator $\mathrm{d}/\mathrm{d}t$ is computed using automatic differentiation and can be thought of as an 'activation operator'. Moreover, the gradients of the loss function are backpropagated through the entire network to train the parameters using a gradient-based optimization algorithm.

Specifically, in this work, we assume that the only observables are noisy data $\{t^n, I^n\}_{n=1}^{M}$ on the number of infected people $I(t)$ that corresponds to the real-world data in Table 1. Moreover, the only initial condition we assume is known as $R(0) = R_0$. Given such data, we are interested in inferring the latent (hidden) quantities $S(t)$ and $R(t)$. The shared parameters of the neural networks for S, I and R in addition to parameters α and β of the differential equations (2) can be learned by minimizing the sum of squared errors loss function

$$\mathrm{SSE} = \sum_{n=1}^{M} |I(t^n) - I^n|^2 + M|R(0) - R_0|^2 + \sum_{i=1}^{4} \sum_{n=1}^{M} |E_i(t^n)|^2. \tag{11}$$

Here, the first term corresponds to the training data on the number of infected $I(t)$ while the last term enforces the structure imposed by Equations (2) and (1) at a finite set of measurement points whose number and locations are taken to be the same as the training data. The middle term corresponds to the initial condition on R. It should be pointed out that the number and locations of the points on which we enforce the set of differential equations could be different from the actual training data. Although not pursued in the current work, this could significantly reduce the required number of training data on I.

The results are reported in Figure 10 where the algorithm learns the parameters α and β to have values 0.454864 and 0.00229365, respectively. Predictions of the model for S, I and
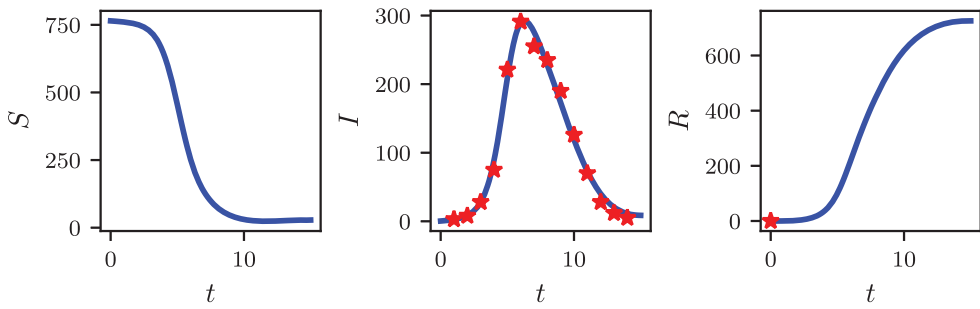
Figure 10. Physics informed deep learning model predictions and comparison to data with $\alpha = 0.454864$ and $\beta = 0.00229365$.

R are plotted using solid blue lines while the data are depicted using red asterisks. Since the total number of training data is relatively small, we chose to optimize the loss function (11) using L-BFGS; a quasi-Newton, full-batch gradient-based optimization algorithm (Liu & Nocedal, 1989). For larger data sets, a more computationally efficient mini-batch setting can be readily employed using stochastic gradient descent and its modern variants (Goodfellow et al., 2016; Kingma & Ba, 2014).

It is worth emphasizing that automatic differentiation is different from, and in several respects superior to, numerical or symbolic differentiation – two commonly encountered techniques of computing derivatives. In its most basic description (Baydin et al., 2018), automatic differentiation relies on the fact that all numerical computations are ultimately compositions of a finite set of elementary operations for which derivatives are known. Combining the derivatives of the constituent operations through the chain rule gives the derivative of the overall composition. This allows accurate evaluation of derivatives at machine precision with ideal asymptotic efficiency and only a small constant factor of overhead. In particular, to compute the required derivatives we rely on Abadi et al. (2016), which is a popular and relatively well-documented open-source software library for automatic differentiation and deep learning computations. In TensorFlow, before a model is run, its computational graph is defined statically rather than dynamically as for instance in PyTorch (Paszke et al., 2017). This is an important feature as it allows us to create and compile the computational graph for the physics informed neural networks (9) only once and keep it fixed throughout the training procedure. This leads to significant reduction in the computational cost of the proposed framework.

4. Statistical inference approaches

Whenever a number of events per time period is observed over time, the time series of counts naturally appear. Moreover, when dealing with count data in a given disease data set such as Table 1, the application of Poisson distribution in modelling such data is appropriate. Looking at the infected numbers in Table 1, and considering the total count of students (i.e. 763), there exists dependency among the infected numbers. Also, as explained earlier, infected students did not get removed from the population until 1 or 2 days after they were infected. Therefore, not only they were contagious and were infecting more students

around them before removal, but they were also counted among the number of infected students during the next days until they were removed.

From the graphs in Figure 3, we now consider the full data, shown in Table 2, as the real-world data set for the statistical modelling of the influenza example. Figure 11 shows how the infected numbers could be related to, and possibly affected by, the convalescent number over time. To account for this potential influence, convalescent count was taken into consideration as a covariate while studying the trend of the infection over time. Both types of dependencies which existed among the observations needed to be addressed. Therefore, a simple count model, where count observations are independent from each other, or a regular longitudinal model, where the same sample is kept for each observation over time, would not be appropriate for modelling this data, estimating the coefficients and making predictions.

A time series model that takes into consideration the dependencies that existed among the counts and adjusts for the multiple counting of some individuals, while adjusting for

Table 2. Full influenza dataset.

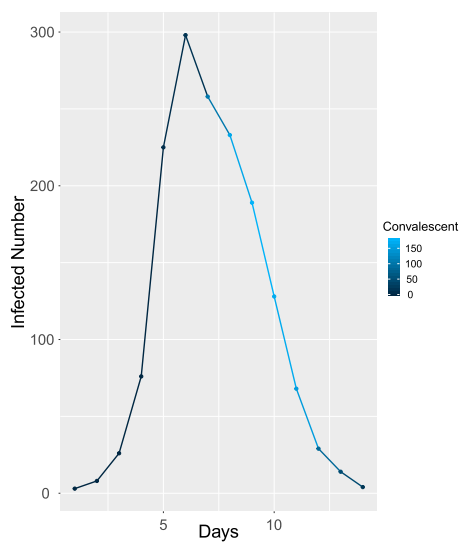| Day	Infected number	Convalescent
1	3	0
2	8	0
3	28	0
4	75	0
5	221	9
6	291	17
7	255	105
8	235	162
9	190	176
10	126	166
11	70	150
12	28	85
13	12	47
14	5	20

Figure 11. Infected and convalescent numbers in the English Boarding School.

other interventions, is fitted to this data. The time series needed to be nonlinear since the measured counts were not following a normal distribution, due to their discrete nature. Many candidates were considered for this data, including autoregressive integrated moving average (ARIMA) models, but only non-linear models could explain the exponential growth of the infected numbers halfway through the study and then the decline in the numbers of infected students on the second half. The nonlinear time series fitted models are shown in Figure 12. The Poisson count regression time series model shown in blue (two dashed line with square) fits the data, and the changing trend of the infected numbers, the best while providing the closest estimates of the number of students infected per day. The reason we opted for this model, which uses a quasi maximum likelihood-based
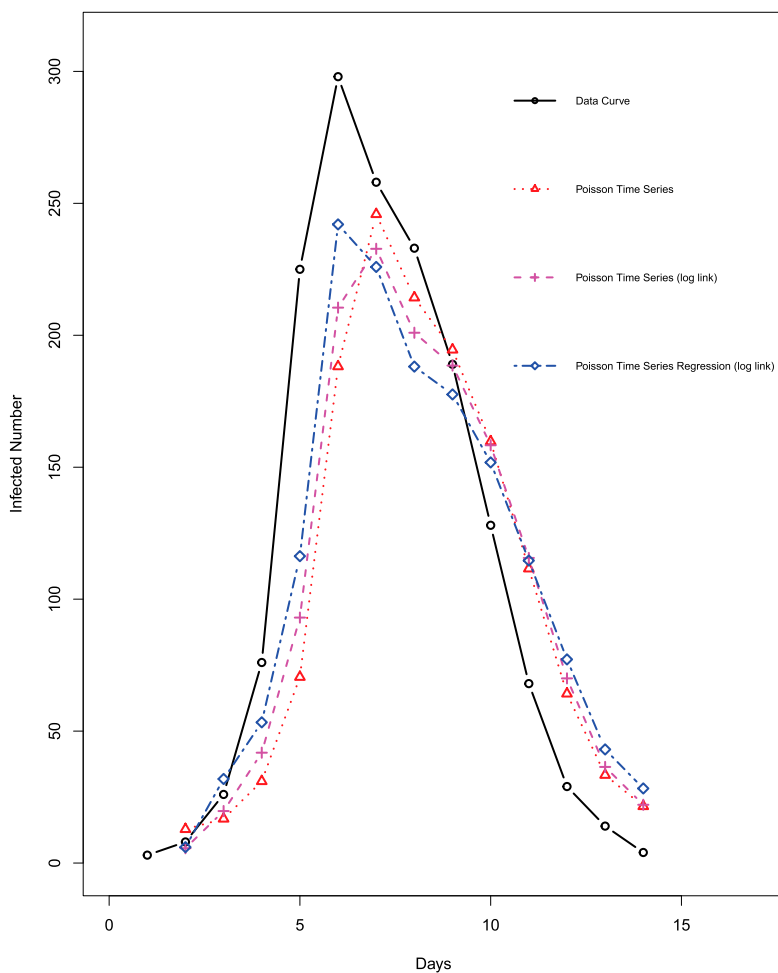

Figure 12. *Statistical models of infectious diseases*: Data (solid black line with circle), Poisson time series (dotted red line with triangle), Poisson time series with logarithmic link (dashed purple line with cross) and Poisson time series regression with logarithmic link (two dashed blue line with square). Non-linear time series regression uses the previous data and covariates while predicting the future values of infections.

technique for estimating the parameters, is to address four issues we were facing while fitting a statistical model to the Influenza data. These issues included small data, two types of dependency among observations which were explained above, and the discrete count nature of the data.

The fitted model within the class of count time series follows generalized linear models and is flexible in describing serial correlation, which can be observed in epidemic disease studies, in a parsimonious way (Liboschik, Kerschke, Fokianos, & Fried, 2016). The special case of this class, called integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) model with its log-linear extension, was applied here to detect the unusual effects of structural changes and different forms of outliers influencing the ordinary pattern of the data. Through conditional likelihood estimation, this model allows modelling such interventions where an intervention directly affects the observation at its occurrence, but not the underlying mean, and then also enters the dynamics of the process. Within this model, we have linked the conditional mean of the observed infection process to its past values, to past observations and to convalescent number as a covariate. We considered a model with the Poisson conditional distribution and the logarithmic link function. The fitted Poisson model using the identity link function is also shown in Figure 12 in red (dotted line with triangle), which did not fit the data as well as when a logarithmic link was used within the count (non-linear) Poisson time series regression. This approach can be applied after the complete time series have been observed, like the Influenza study, to build a general model if for instance this infection happens again within the same or a different population. The model can also be built during an ongoing outbreak to estimate the growth rate and the future counts.

The formulation of this model class is explained next, based on the notations used in Liboschik, Fokianos, and Fried (2017). Let Y_t be a count time series (infected number here) and X_t be a covariate vector (convalescent counts here). Assume $\mathcal{F}_t$ is the history of the joint process of the count series and covariates $\{Y_t, \lambda_t, X_{t+1}\}$ up to time t including the covariate information at time $t + 1$, where $\lambda_t = \mathbb{E}(Y_t|\mathcal{F}_{t-1})$ is the rate at which the Poisson process happens and the mean of counts at time t conditioned on the joint process of the previous counts and covariates. The general model is of the form

$$g(\lambda_t) = \beta_0 + \sum_{k=1}^{p} \beta_k \widetilde{g}(Y_{t-i_k}) + \sum_{l=1}^{q} \alpha_i g(\lambda_{t-j_l}) + \eta' X_t, \tag{12}$$

where g is a link function (log link here), $\widetilde{g}$ is a transformation function, η is the parameter vector which corresponds to the effects of covariates, and $P = i_1, \ldots, i_p, Q = j_1, \ldots, j_q$ are sets of integers. These models are called INGARCH model of order p and q and also as autoregressive conditional Poisson model. A distribution assumption on the conditional distribution of Y_t allows specification of the Poisson distribution for the count data implemented in this study.

The use of quasi maximum likelihood-based inferences for such models allows making 1-step-ahead predictions (Liboschik et al., 2016). After applying a log-linear model with the logarithmic link, as it allows for negative covariate effects, and adding a first-order autoregressive term, to capture the short range serial dependence, a Poisson distribution was shown to be sufficient. Probability integral transform histograms, a marginal calibration plot and the scoring rules (not shown here) were checked for finalizing the

Table 3. Fitted time series.

Parameters	Estimate	Std. Error	CI (lower)	CI (upper)
(Intercept)	2.001	0.236	1.539	2.464
β_1	0.892	0.056	0.782	1.003
α_1	−0.279	0.068	−0.412	−0.146
η_1	−0.0013	0.0006	−0.0024	−0.0001

distribution. Models with and without covarying convalescent counts were fit and as obvious from Figure 12, the model excluding the covariate (the dashed purple curve with cross) was not as good as the model that included this covariate (two dashed blue curve with square). Table 3 shows the results of fitting the preferred model to the Influenza data.

Using the general INGARCH time series model, Equation (12), the fitted model for the number of infected individuals Y_t in day t is given by $Y_t|\mathcal{F}_{t-1}$ Poisson(λ_t) with

$$\log \lambda_t = 2 + 0.892 Y_{t-1} - 0.279 \lambda_{t-1} - 0.001 X_t, \quad t = 1, \ldots, 14,$$

where X_t is the covariate, at time t, and Y_{t-1} and λ_{t-1} are the previous observations and their lagged conditional means, respectively. As seen from the confidence intervals of the estimated parameters, all the parameters of this model are significantly influencing the number of infections. There is a notable dependence to the number of infected individuals of the preceding day. This model could be fitted to any infectious disease data set, to predict number of infections and recoveries. As seen from the results, the predicted values are close to the actual number of infections and the course for the disease.

5. Application of methods to extended dataset

In this section, we discuss the application of the parameter estimation, deep learning and statistical approaches discussed in the earlier sections applied to an extended dataset with over 100 observations as a proof of concept. We want to show through this example that the methods presented in the earlier sections work consistently for a larger dataset as well. The new extended data illustrated in Figure 13 was created from the original Influenza data by interpolating the 14-day dataset to every 2 h during the 14 days. The reason for doing this was to evaluate the performance of the aforementioned models with larger data. The methods applied worked well with the new larger dataset with 157 observations and the detailed results are included within the next three sections.

5.1. Estimating optimal parameters via parameter estimation

First we considered the parameter estimation technique using the unconstrained nonlinear optimization that uses a regression approach to search for the best parameters by minimizing an objective function represented by the differences in the 2-h counts from observed data and the computer simulated data. Using initial conditions $S_0 = 762, I_0 = 1, R_0 = 0$ with a poor guess for $\alpha_{\text{guess}} = 0.5, \beta_{\text{guess}} = 0.01$, the optimization algorithm estimates the parameters to be $\beta = 0.0002$ and $\alpha = 0.0401$. The predicted dynamics of the three subpopulations for these parameter values along with comparison to the new extended dataset is illustrated in Figure 14.
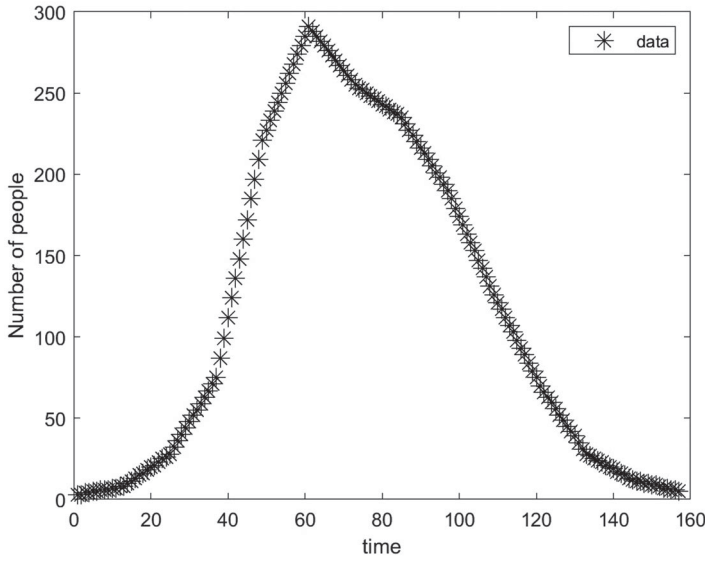
Figure 13. Extended Dataset interpolated every 2 h from the 14-day Influenza Dataset.
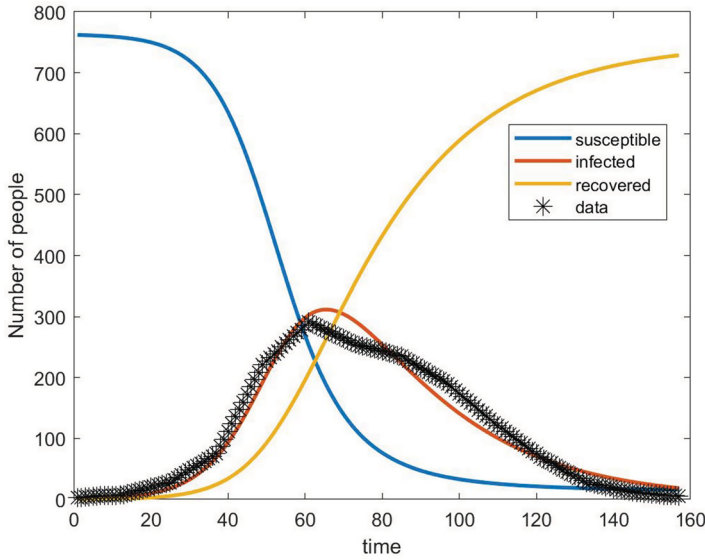

Figure 14. Dynamics of the SIR model predictions and comparison to data for $\alpha = 0.0401$ and $\beta = 0.0002$.

5.2. Estimating optimal parameters via deep learning

Next, we considered applying the Physics informed Deep Learning approach described in this work to the extended dataset in Figure 13. As before, we acquire the required derivatives to compute the residual networks E_1, E_2, E_3 and E_4 and employ densely connected physics uninformed neural network with 1 hidden layer and 32 neurons per hidden layer which takes the input variable t and outputs S, I and R. The shared parameters of the neural
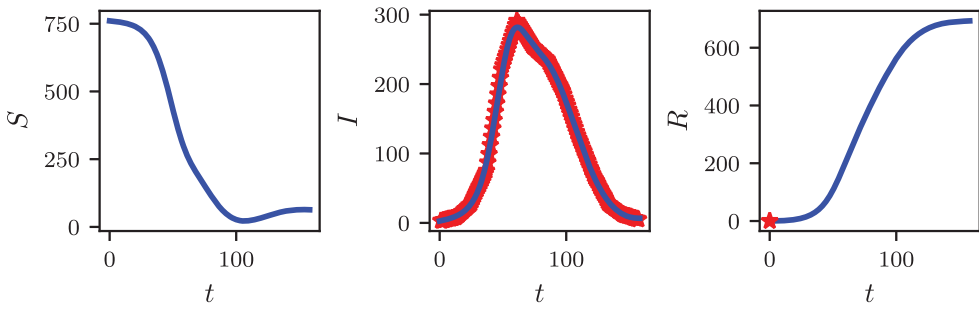
Figure 15. Physics informed deep learning model predictions and comparison to data with $\alpha = 0.0374007$ and $\beta = 0.00018133$.

networks for S, I and R in addition to parameters α and β of the differential equations (2) are then learned using Deep Learning approaches described by minimizing the sum of squared errors loss function (11). The results are reported in Figure 15 where the algorithm learns the parameters α and β to have values 0.0374007 and 0.00018133.

5.3. Statistical inference approaches for extended dataset

To the best of our knowledge, literature of the modern time series approaches used in this paper does not report a large sample size requirement for the INGARCH models (Liboschik et al., 2016, Liboschik, Fokianos, & Fried 2017). As seen in Section 4, they performed well while modelling small data. However, due to the fact that the variance of the estimated coefficients can be slightly higher due to the smaller size of the original data used in this study, we applied the same techniques to a larger data set and compared the results.

Same time series approaches that enabled us to take into consideration the dependency among the infected counts while adjusting for the multiple counting of some individuals and other interventions are fitted to the larger data. The time series needed to be nonlinear since the discrete measured counts were not following a normal distribution. Similar to Section 4, many candidates were considered for this data, including ARIMA models, but only non-linear count Poisson time series models could explain the exponential growth of the infected numbers for the first half of the study and the immediate decline in the numbers of infected students on the second half of the data. The nonlinear time series fitted models to the larger data are shown in Figure 16.

The count Poisson regression time series model with logarithmic link function, which included convalescent numbers as a covariate, had the best fit when working with the smaller data. The same model applied to the lager data resulted in the fitted values shown in blue (solid line with asterisk) in Figure 16. Clearly, this model fits the data, and the changing trend of the infected numbers, well and even better than when the smaller data were used. For the larger data, the fit of this model was not the best among the fitted models and we anticipate the fit of this model to improve for different data sets and scenarios by the inclusion of more covariates. The same count (non-linear) Poisson time series regression fitted model using the identity link function is also shown in Figure 16 in red (solid line with triangle), which has the best fit. The large data characteristics of the current data is helping with this fit. The same model excluding the covariate is also shown in Figure 16 in
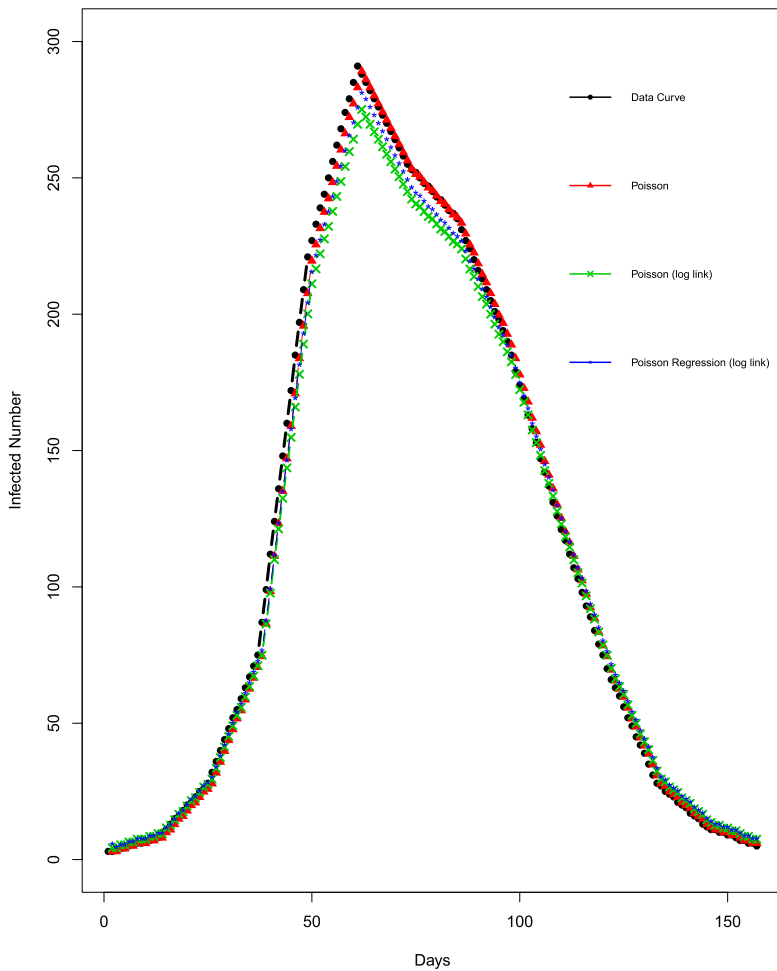
Figure 16. *Statistical models of infectious diseases*: Data (solid black line with circle), Poisson time series (solid red line with triangle), Poisson time series with logarithmic link (solid green line with cross) and Poisson time series regression with logarithmic link (solid blue line with asterisk).

green (solid line with cross), which has a good fit but not as good as the previous model. Due to the high number of observations, it might be difficult to see the differences among the fitted models in Figure 16. Therefore, a zoomed-in version of this figure, where the highest difference among fitted values was present, is provided in Figure 17 to show the small differences in the fit of the models better.

After visually comparing the fit of the time series models through Figures 16 and 17, we show three types of fit indices in Table 4 to numerically compare the fit of the afore-mentioned models. Akaike information criterion (AIC), Bayesian information criterion (BIC) and quasi information criterion (QIC) are the fit indices reported here. They represent the relative quality of a statistical model by measuring the amount of information lost by that model. These criteria estimate the quality of each model, relative to other models and therefore cannot be interpreted based on their absolute values. They rather need to be compared across models and the model with the smallest AIC, BIC and QIC has the best
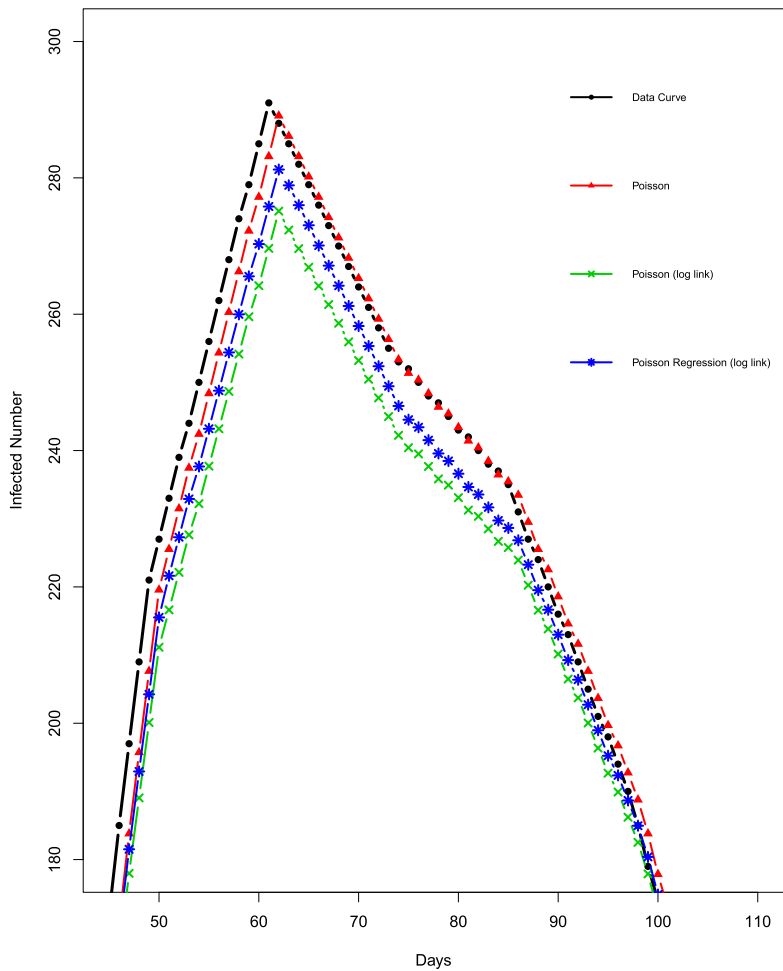
Figure 17. *Statistical models of infectious diseases* (*Zoomed-in version of Figure* 16): Data (solid black line with circle), Poisson time series (solid red line with triangle), Poisson time series with logarithmic link (solid green line with cross) and Poisson time series regression with logarithmic link (solid blue line with asterisk).

Table 4. Fit indices of time series models.

Fit Index	Poisson	Poisson (log link)	Poisson regression (log link)
AIC	977.15	1096.68	1205.22
BIC	983.26	1105.85	1217.45
QIC	977.22	1166.92	1531.44

fit among the comparable fitted models. Table 4 shows that the first model, which is count Poisson time series regression with identity link function, has the best fit. This result was obvious in Figures 16 and 17 as well. Note that the other two models, which were count Poisson time series regression with logarithmic link function without and with a covariate, do not have a big difference in terms of model fit; therefore, they each could be a reasonable option to be adopted by researchers for such scenarios.

Table 5. Forecast accuracy of time series models.

Accuracy Index	Poisson	Poisson (log link)	Poisson regression (log link)
RMSE	4.82	9.29	10.85
MAE	3.39	6.45	5.65
MAPE	5.51	18.80	32.87

Evaluating the forecast accuracy of the fitted models is also important within time series approaches. Table 5 shows the comparisons of three measures of forecast accuracy. These accuracy measures are root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). Interpreting these measures are slightly different. A forecast method that minimizes the MAE will lead to forecasts of median, while a forecast method that minimizes the RMSE will lead to forecasts of the mean. Like the fit indices, these forecast accuracy measures need to be compared across models and the model with the smallest RMSE, MAE and MAPE is the most accurate in terms of forecasting. Table 5 shows that the first model, which is count Poisson time series regression with identity link function, has the best forecast accuracy. The other two models, which were count Poisson time series regression with logarithmic link function without and with a covariate, compete with each other for the next best model in terms of forecast accuracy but do not have big differences; therefore, either one of them could be a reasonable option to be adopted by researchers for forecasting.

As it can be seen from the results of the statistical models, a larger sample size improved the fit of the models. As it is clear from the results discussed above, the predicted values are very close to the actual number of infections and the course for the disease, which indicates these statistical inference methods can be adopted by researchers for varying sizes of samples in the spread of such infectious diseases.

6. Discussion and conclusion

Since the introduction of the SIR model about a century ago, there has been a lot of work in using variations of the model to understand and track the spread of infectious diseases. The compartmental SIR model that consisted of a system of coupled ordinary differential equations gave a framework to predict how a disease spreads, for example the prevalence (total number of infected) or the duration of an epidemic. Also, the model allowed for understanding how different factors may affect the outcome of the epidemic which can influence policy. While there has been a lot of progress made in utilizing SIR (and its variants) to understand propagation of disease, there is still a great need to develop novel techniques to estimate important model parameters such as transmission and recovery rates that help determine the nature of the disease propagation.

We started by reviewing some of the analytical and heuristic methods that employ final size computation formulas or Berkeley Madonna method to determine the parameters and then using them to compare the prediction to the actual real-world data. It was clear that there are variations in these comparisons leading us to believe that there is still scope to identify algorithms to determine best-fit parameters for a given data set. In this regard, the simulation optimization approach was presented which employed a Nelder–Meade based algorithm and was able to do a better job at estimating optimal parameters than many of the existing heuristic methods presented earlier.

Following this, the focus of the paper was on two main contributions including the application of physics based deep learning for infectious diseases and statistical inference-based time series methods. The authors believe that both these are introduced for the first time to the mathematical biology community to help understand the need to develop less expensive, more optimal computational and statistical methods that can provide reliable information to predict disease dynamics with minimal data. The performance of these new approaches was illustrated for a real-world benchmark application. To support the performance of the methods, we applied the techniques to both small and large data sets. We were able to show that in both cases, the parameter estimation approach and the Physics-based deep learning approaches were able to recover consistent values of the *transmission* and *recovery* rates. Statistical time series approaches used within this study also performed well and resulted in a good fit and forecast accuracy in both cases. The fit and forecast accuracy improved by the increase in the size of the data. Our results clearly indicate that these approaches are reliable candidates for parameter estimation and prediction of the spread of a disease. The goal of this work was to introduce these approaches and hence they were applied to a simple SIR model that has been used to track the spread of Influenza. We hope to consider the application of these for more advanced models that have been established to study the spread the propagation of other diseases such as Zika (Padmanabhan, Seshaiyer, & Castillo-Chavez, 2017) in a forthcoming paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . , Zhang, X. (2016, November). Tensorflow: A system for large-scale machine learning. *CoRR*. Retrieved from http://arxiv.org/abs/1605.08695.

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control* (Vol. 15). Oxford: Oxford University Press.

Anonymous (1978). Influenza in a boarding school. *British Medical Journal*, *1*, 578. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1603269/pdf/brmedj00115-0064.pdf

Backer, J., Berto, A., McCreary, C., & Martelli, F. (2012). Transmission dynamics of hepatitis e virus in pigs: Estimation from field data and effect of vaccination. *Epidemics*, *4*, 86–92.

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: A survey. *Journal of Marchine Learning Research*, *18*, 1–43.

Blower, S. M., & Dowlatabadi, H. (1994). Sensitivity and uncertainty analysis of complex models of disease transmission: An HIV model, as an example. *International Statistical Review/Revue Internationale de Statistique*, *62*(2), 229–243.

Brauer, F., & Castillo-Chavez, C. (2012). *Mathematical models in population biology and epidemiology* (Vol. 40). New York: Springer.

Caley, P., & Ramsey, D. (2001). Estimating disease transmission in wildlife, with emphasis on leptospirosis and bovine tuberculosis in possums, and effects of fertility control. *The Journal of Applied Ecology*, *38*, 1362–1370.

Capaldi, A., Behrend, S., Berman, B., Smith, J., Wright, J., & Lloyd, A. L. (2012). Parameter estimation and uncertainty quantication for an epidemic model. *Mathematical Biosciences and Engineering*, *9*(3), 553–576.

Chitnis, N., Hyman, J. M., & Cushing, J. M. (2008). Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. *Bulletin of Mathematical Biology*, *70*(5), 1272.

Elkadry, A. (2013). *Transmission rate in partial differential equation in epidemic models* (Ph.D. thesis). Huntington, West Virginia: Marshal University.

Finkenstädt, B. F., & Grenfell, B. T. (2000). Time series modelling of childhood diseases: A dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(2), 187–205.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). Cambridge: MIT Press.

Hadeler, K. P. (2011). Parameter identification in epidemic models. *Mathematical Biosciences*, *229*(2), 185–189.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In *Neural networks for perception* (pp. 65–93). Orlando, FL: Harcourt Brace & Co.

Huang, Y., Liu, D., & Wu, H. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics*, *62*(2), 413–423.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *115*(772), 700–721.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. Preprint arXiv:1412.6980.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436.

Liboschik, T., Fokianos, K. and Fried, R. (2017). tscount: An R package for analysis of count time series following generalized linear models. *Journal of Statistical Software*, *82*(5), 1–51.

Liboschik, T., Kerschke, P., Fokianos, K., & Fried, R. (2016). Modelling interventions in INGARCH processes. *International Journal of Computer Mathematics*, *93*(4), 640–657.

Liu, D. C., & Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, *45*(1–3), 503–528.

Longini, I. M., Koopman, J. S., Haber, M., & Cotsonis, G. A. (1988). Statistical inference for infectious diseases: Risk-specific household and community transmission parameters. *American Journal of Epidemiology*, *128*(4), 845–859.

Macey, R., Oster, G., & Zahnley, T. (2000). *Berkeley Madonna user's guide*. Berkeley, CA: University of California.

Martcheva, M. (2015). *An introduction to mathematical epidemiology* (Vol. 61). New York: Springer.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239–245.

Murray, J. D. (1989). *Mathematical Biology. Biomathematics Vol 19*. New York: Springer-Verlag.

Nsoesie, E. O., Beckman, R. J., Shashaani, S., Nagaraj, K. S., & Marathe, M. V. (2013). A simulation optimization approach to epidemic forecasting. *PLoS ONE*, *8*(6), e67164.

O'Dea, E. B., Pepin, K. M., Lopman, B. A., & Wilke, C. O. (2014). Fitting outbreak models to data from many small norovirus outbreaks. *Epidemics*, *6*, 18–29.

Padmanabhan, P., Seshaiyer, P., & Castillo-Chavez, C. (2017). Mathematical modeling, analysis and simulation of the spread of Zika with influence of sexual transmission and preventive measures. *Letters in Biomathematics*, *4*(1), 148–166.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., . . . , Lerer, A. (2017). Automatic differentiation in PyTorch. Retrieved from https://openreview.net/forum?id = BJJsrmfCZ.

Pollicott, M., Wang, H., & Weiss, H. (2012). Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem. *Journal of Biological Dynamics*, *6*(2), 509–523.

Raissi, M., & Karniadakis, G. E. (2018). Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, *357*, 125–141.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017a). Inferring solutions of differential equations using noisy multi-fidelity data. *Journal of Computational Physics*, *335*, 736–746.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017b). Machine learning of linear differential equations using Gaussian processes. *Journal of Computational Physics*, *348*, 683–693.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017c). Physics informed deep learning (Part I): Data-driven solutions of nonlinear partial differential equations. Preprint arXiv:1711.10561.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017d). Physics informed deep learning (Part II): Data-driven discovery of nonlinear partial differential equations. Preprint arXiv:1711.10566.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2018). Multistep neural networks for data-driven discovery of nonlinear dynamical systems. Preprint arXiv:1801.01236.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Wallinga, J., Lávy-Bruhl, D., Gay, N. J., & Wachmann, C. H. (2001). Estimation of measles reproduction ratios and prospects for elimination of measles by vaccination in some Western European countries. *Epidemiology and Infection*, *127*(2), 281–295.

Yong, K. E., Mubayi, A., & Kribs, C. M. (2015). Agent-based mathematical modeling as a tool for estimating Trypanosoma cruzi vector–host contact rates. *Acta Tropica*, *151*, 21–31.