



PAPER

# Predicting COVID-19 pandemic waves with biologically and behaviorally informed universal differential equations

Bruce Kuwahara<sup>a</sup> and Chris T. Bauch<sup>a,\*</sup>

<sup>a</sup>Department of Applied Mathematics, University of Waterloo, 200 University Ave W, N2L 3G1, Ontario, Canada

\*To whom correspondence should be addressed: [cbauch@uwaterloo.ca](mailto:cbauch@uwaterloo.ca)

FOR PUBLISHER ONLY Received on Date Month Year; accepted on Date Month Year

## Abstract

In the early stages of the COVID-19 pandemic, it became clear that pandemic waves and population responses were locked in a mutual feedback loop. The initial lull following strict interventions in the first wave often led to a second wave, as restrictions were relaxed. We test the ability of new hybrid machine learning techniques, namely universal differential equations (UDEs) with learning biases, to make predictions in a such a dynamic behavior-disease setting. We develop a UDE model for COVID-19 and test it both with and without learning biases describing simple assumptions about disease transmission and population response. Our results show that UDEs, particularly when supplied with learning biases, are capable of learning coupled behavior-disease dynamics and predicting second waves in a variety of populations. The model predicts a second wave of infections 55% of the time across all populations, having been trained only on the first wave. The predicted second wave is larger than the first. Without learning biases, model predictions are hampered: the unbiased model predicts a second wave only 25% of the time, typically smaller than the first. The biased model consistently predicts the expected increase in the transmission rate with rising mobility, whereas the unbiased model predicts a decrease in mobility as often as a continued increase. The biased model also achieves better accuracy on its training data thanks to fewer and less severely divergent trajectories. These results indicate that biologically informed machine learning can generate qualitatively correct mid to long-term predictions of COVID-19 pandemic waves.

**Key words:** machine learning, compartmental model, COVID-19, physics informed neural networks

### Significance statement

Universal differential equations are a relatively new modelling technique where neural networks use data to learn unknown components of a dynamical system. We demonstrate for the first time that this technique is able to extract valuable information from data on a coupled behaviour-disease system. Our model was able to learn the interplay between COVID-19 infections and time spent travelling to retail and recreation locations in order to predict a second wave of cases, having been trained only on the first wave. We also demonstrate that adding additional terms to the universal differential equation's loss function that penalize implausible solutions improves training time and leads to improved predictions.

## Introduction

The COVID-19 pandemic generated an enormous demand for mathematical models to predict cases and guide policy [1, 2, 3]. These models were often mechanistic in nature, seeking to represent known or hypothesized transmission mechanisms in

a mathematical, stochastic, or agent-based framework [4, 5, 6]. The advance warning provided by these models allowed public health institutions to prepare by implementing policies to mitigate the second wave when it arrived [7, 8]. Modelling efforts were widely applied to investigate the impact of public

health measures such as testing [9], school and workplace closures [10, 11], vaccination strategies [12, 13, 14] or to stimulate action by projecting the impacts of a worst case ‘do nothing’ scenario where governments and populations did not attempt to mitigate the pandemic [8, 10].

Fortunately, most governments and members of the public did respond to the pandemic, by taking measures to reduce case incidence. Numerous studies have shown that non-pharmaceutical interventions such as lockdowns, school closures, and social distancing protocols reduce case notifications and health impacts of COVID-19 [15, 16, 17, 18]. The anticipation of infection risk in the face of rising case incidence supports adherence to these measures [19]. However, as risk of infection wanes, so too does willingness to abide by these mentally fatiguing [20, 21] and economically costly measures [21]. The relaxation of mitigation efforts may then potentially result in another pandemic wave. This two-way interaction—where infection spread influences behaviour, which influence infection spread, in turn—suggests the concept of coupled behavior-disease systems [22] may be useful for studying COVID-19 pandemic waves.

As such, many mechanistic models informed by economic, social or psychological assumptions have incorporated behaviour-disease dynamics to study the impact of interventions in the context of population behavioural feedbacks [23, 24, 25, 26, 27, 28, 29], including for COVID-19 [30, 12, 31, 32, 33]. Among the most valuable insights provided by these models is the occurrence of multiple pandemic waves, which are predicted under a wide range of conditions by these models due to waning stringency causing a resurgence of the infection [34, 12, 35, 36]. With hindsight, we can confirm that these models were correct – second waves occurred virtually everywhere during the COVID-19 pandemic (and before the arrival of new variants).

Alongside these mechanistic models, the plethora of epidemiological, sociological, mobility, and economic data generated by the pandemic allowed machine learning models to flourish [37, 38, 39, 40, 41]. These models have proven adept at integrating vast quantities of data on a multitude of factors (including behaviour) affecting disease spread. Consequently, they often adapt better to regional variability compared to mechanistic models [41, 40]. However, machine learning models have significant drawbacks. They are able to fit existing data well and accurately predict days to a couple of weeks into the future, but pay for this predictive accuracy with reduced interpretability compared to traditional models [37]. Compared to mechanistic models with relatively few easily-understood parameters, it is far more difficult to extract qualitative understandings of disease dynamics (such as second waves) from the hundreds or thousands parameters in purely machine-learning models. They are also easy to over-fit (although mechanistic models also suffer from this risk), meaning their predictive value may be limited.

Recently, advances in high-performance automatic differentiation have enabled new techniques that combine the interpretability and qualitative understanding from mechanistic models with the potentially higher predictive power and scalability of machine learning. Physics-informed machine learning (PIML) is one such methodology. The key idea is to create ML models that encode physical laws by inferring them from large amounts of data (observational bias), building them into the model’s architecture (structural bias), or training the model to uphold them (learning bias) [42]. Of particular interest for qualitative epidemic modelling are the latter two biases, as they reduce the model’s reliance on large amounts of data. By

incorporating these biases, the model is prevented (in the case of structural biases) or at least discouraged (for learning biases) from making biologically impossible predictions. Learning biases can also discourage overfitting the data by introducing other objectives for the model.

The majority of learning bias applications thus far have been solving various forms of partial differential equations (PDEs) [43, 44, 45]. In these models, a neural network is trained to simultaneously fit data and to satisfy a PDE. In addition to physics, learning biases have been used in biologically-informed machine learning (BIML) applications. These include blood flow dynamics [46], drug responses [47], and cancer detection and classification [48, 49, 50]. This approach has yet to be applied to COVID-19 modeling, but has seen great success for solving large and complex partial differential equations.

A novel technique for structural bias is universal differential equations (UDEs). This method involves training neural networks embedded in differential equation models. Known dynamics can be included explicitly, while leaving unknown processes to be learned by the neural network [51]. The explicit parts of the UDE can be made to retain valuable laws such as invariant quantities. UDEs have been applied successfully on predator-prey models, metabolic networks, batteries, and photonics [51, 52]. For instance, recent research uses a neural network to learn the change in COVID-19 quarantine measures in a population over time, within the framework of a modified QSEIR (quarantine, susceptible, infectious, recovered) model. The trained network was then used to quantify the effectiveness of those measures for different regions [39, 53].

The ability of hybrid machine learning approaches – such as UDEs and PIML – to make qualitative long-term predictions about epidemic dynamics of the sort provided by mechanistic models has not yet been widely tested. This motivated our research: our objective was to combine observational biases (UDEs) with satisfiable learning biases in a coupled behaviour-disease model for COVID-19. We trained a compartmental UDE model to fit behavioural and epidemic data while penalizing deviations from several simple socio-biological assumptions. We hypothesized that a UDE model can learn the pattern of behaviour-disease interactions and hence predict a second wave (either qualitatively or quantitatively), having only seen the first wave (and its learning biases). We also hypothesized that without those learning biases, the model will learn much less effectively.

## Results

A complicated mathematical model can easily be made to fit an epidemic curve, but runs the risk of over-fitting the data and thus not being useful for prediction [54, 55]. Simpler mathematical models allow us to test our hypotheses by incorporating aspects we understand without becoming overburdened by details that we cannot reliably describe mathematically [55].

Hence, we used a UDE framework that allows us to leave the behaviour-disease dynamics of a simple compartmental behaviour-disease model unspecified, save for a few plausible assumptions (“learning biases”). In doing so, we can test the validity of those assumptions. Compartmental models divide the human population into mutually exclusive compartments based on infectious status, and which are generally implemented as differential equations. These models

have been a mainstay of mathematical epidemiology for decades [5].

The algorithm learned the manner in which force of infection responds to mobility and manner in which mobility responds to its current value, the number of active cases, recent new cases, and recovered cases. The learning biases inform the model with several plausible assumptions: namely that force of infection increases with mobility, that mobility decreases with more active and recent cases, that mobility tends toward 0 (the pre-COVID average) in the absence of cases, that this tendency is stronger the more people have recovered, and that mobility cannot fall below a 100% reduction or exceed a 200% increase from the pre-COVID average (see Methods). The learning biases strongly discourage infeasible values of mobility and make data-fitting relatively less important for the optimizer. As a result, the model makes out-of-sample predictions (i.e. second waves) frequently.

In order to ensure consistent and repeatable results, we ran the model on each region 100 times both with and without learning biases. We trained the algorithm on the first wave and tested whether it could predict the second wave. Overall, the model with learning biases was successful in every region in which we tested it, though some more so than others. It consistently learned to fit the data and constraint losses, predicted second waves, and seldom made biologically implausible predictions.

We compared UDE models with and without learning bias. The model without learning biases, while not entirely a failure, was much less successful. Though it was generally able to fit the data, it predicted second waves much less frequently and made many more unrealistic predictions. Details are provided in the following subsections.

### Model Predictions

To get a sense of the model's average behaviour, we plotted the median prediction of the 100 simulations for each region. An example for New York can be found in Section 2, Figure 2 (analogs for other regions can be found in the Supplementary Appendix, figures 3-13). The model with learning biases has consistent behaviour within the training region. The median prediction shows a small second wave and the inter-quartile range shows one of similar size to the first. The model without learning biases fits the data comparably well, but has greatly reduced variability outside of the training region. Second wave predictions are smaller or non-existent, typically only suggested by the upper quartile rather than the median. Section 2.1 Table 1 gives a numerical summary of the model's performance across all regions. Section 2.1 Figure 1 shows a graphical summary of model performance across all regions.

### Biological Feasibility

Both models, with and without learning biases, tended to make feasible predictions, in the sense that all model states remained within their respective bounds. The biased model was stable 88% of the time, while the unbiased model was stable 85% of the time, across all regions.

However, when evaluating the learning bias loss functions on the trained models, it becomes clear that the model with learning biases is more reliable in this regard. The biased model achieves better losses across all loss objectives, including accuracy, compared to the unbiased model. Comparison of all loss functions can be found in the supplementary material.

The unbiased model does particularly poorly on the mobility upper and lower bounds (on the order of  $10^4$  times worse than biased), and tendency for mobility to return to baseline in the absence of infection (roughly  $10^3$  times worse).

### Second wave prediction

As a more robust metric for second wave prediction, we counted the number of local maxima exceeding at least  $10^{-3}$  in the infected time series for each model simulation. With learning biases, the model predicted second waves regularly for most regions.

The unbiased model, meanwhile, rarely predicted second waves (Section 2.1, Table 1). Its best performance was on Quebec, where it predicted second waves 51% of the time. This was also the only region in which it outperformed the biased model, which predicted second waves 48% for of the time. Otherwise, it predicted second waves less than 66% as often as the biased model, sometimes as little as 1.6% as often. It predicted no second waves at all for British Columbia.

Most of the time, both models predict the second wave too early (exceptions being Texas and Quebec), but the biased model's estimate is usually closer (only excepting Texas and Quebec). In terms of wave size, the biased model typically overestimates while the unbiased model shows no clear pattern.

### Transmissibility

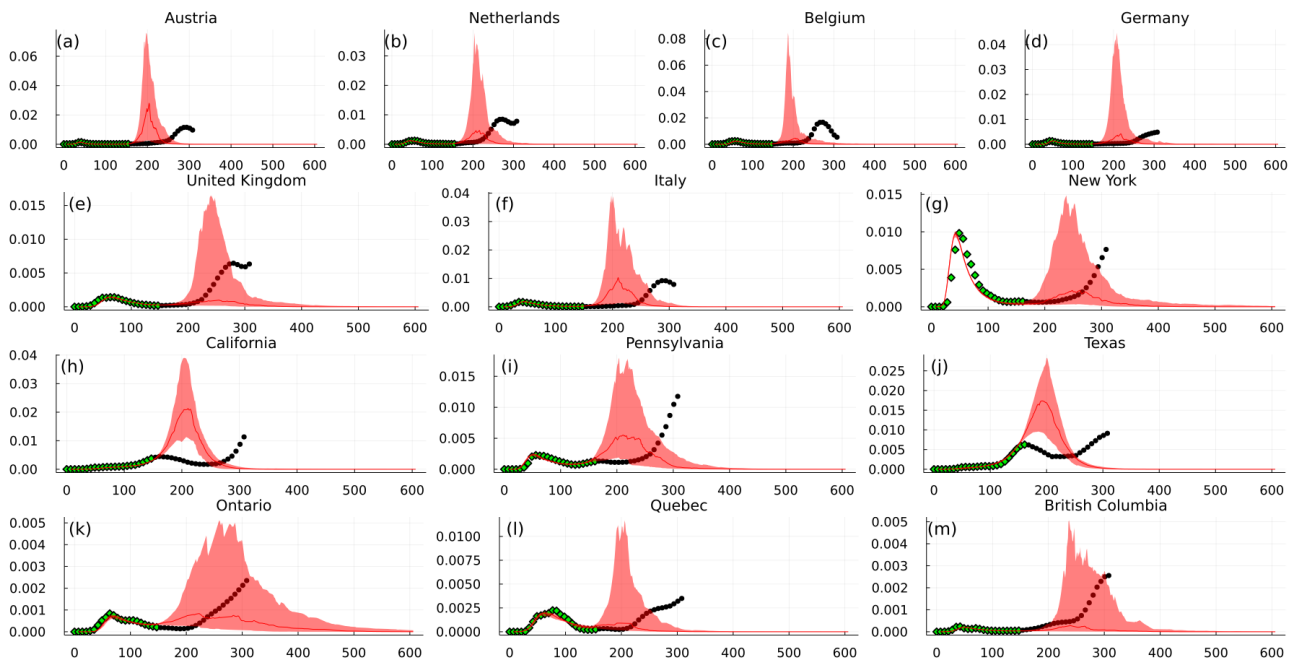
One of the main uses of this model is that the trained neural network representing the force of infection can, once trained, be analyzed to examine the learned relationship between mobility and the transmission rate.

Section 2.4 Figure 3 shows the distribution in the response of  $\beta$  to mobility predicted by the model with learning biases for New York. The models all converge on the same relationship within the training region and on low out-of-sample values, but they diverge for large ones. It is also worth noting that the prediction is, as expected, monotonically increasing. Once again, all regions demonstrate similar behaviour (see Supplementary Appendix figures 14-24).

As with the time series predictions, the model without learning biases fits the data about equally well within the region on which it has been trained. However, outside that region, it extrapolates a flatter curve that is about equally likely to be higher or lower than the median.

For a quantitative sense of how  $\beta$  responds to mobility, we evaluated each trained network at the baseline value of mobility to determine the value of  $\beta$ , and hence  $R_0 (= \beta/\gamma)$ , the basic reproduction number of the virus at the baseline value of mobility. We also applied Newton's method to the trained neural network to find the value of mobility ( $M_{crit}$ ) at which  $R_0$  drops below 1, the value below which the infection will die out [56]. Results for the biased model can be found in table Section 2.4 Table 2. The unbiased model results are negligibly different for the  $R_0$ . The results for  $M_{crit}$  are more variable. These unbiased model results can be found in the Supplementary Appendix Table 1.

The  $R_0$  predictions, averaged over all simulations for a given region, range from 1.60 (British Columbia) to 2.60 (Germany). While estimates of  $R_0$  for COVID-19 vary significantly between countries and times, this is in line with estimates of between 2.4 and 2.4 for the original COVID-19 strain [57, 58, 59, 60]. It is also consistent with other models, which found Germany and the Netherlands to have higher values [61].



**Fig. 1.** Infection prevalence time series predictions for all regions produced by the model with learning biases. Infection prevalence is the proportion of the total population that is infected at any given time. Green dots represent training data (first 22 weeks) and black dots show unseen data (a further 23 weeks). Predictions are generated using the median (solid line) and interquartile range (ribbon) of 100 independently-trained instances of the model per region.

**Table 1.** Summary of second wave predictions: Listed as biased model / unbiased model

Region	True wave time <sup>1,2</sup>	True wave size <sup>1</sup>	2nd wave rate	Mean wave timing <sup>1,3</sup>	Mean wave size
Austria	294	$1.15 \times 10^{-2}$	0.86 / 0.50	222 / 197	$5.88 \times 10^{-2}$ / $2.67 \times 10^{-2}$
Belgium	273	$1.67 \times 10^{-2}$	0.68 / 0.23	217 / 185	$5.18 \times 10^{-2}$ / $5.71 \times 10^{-3}$
Germany	315	$4.90 \times 10^{-3}$	0.78 / 0.45	259 / 244	$4.53 \times 10^{-2}$ / $2.87 \times 10^{-2}$
Netherlands	273	$8.50 \times 10^{-3}$	0.68 / 0.23	226 / 119	$4.09 \times 10^{-2}$ / $2.43 \times 10^{-3}$
Italy	287	$9.18 \times 10^{-3}$	0.83 / 0.27	214 / 151	$4.46 \times 10^{-2}$ / $6.47 \times 10^{-3}$
United Kingdom	280	$6.44 \times 10^{-3}$	0.63 / 0.40	266 / 140	$2.95 \times 10^{-2}$ / $1.16 \times 10^{-3}$
California, USA	336	$1.85 \times 10^{-2}$	0.17 / 0.010	454 / 228	$2.74 \times 10^{-3}$ / $2.91 \times 10^{-4}$
New York, USA	350	$1.26 \times 10^{-2}$	0.79 / 0.39	276 / 190	$1.94 \times 10^{-2}$ / $6.42 \times 10^{-3}$
Pennsylvania, USA	329	$1.29 \times 10^{-2}$	0.80 / 0.19	212 / 93	$2.05 \times 10^{-2}$ / $1.38 \times 10^{-3}$
Texas, USA	217	$3.27 \times 10^{-3}$	0.25 / 0.021	342 / 183	$9.89 \times 10^{-3}$ / $2.45 \times 10^{-4}$
British Columbia, Canada	308	$2.56 \times 10^{-3}$	0.047 / 0	357 / NA	$2.52 \times 10^{-3}$ / 0
Ontario, Canada	343	$3.54 \times 10^{-3}$	0.098 / 0.29	158 / 300	$4.19 \times 10^{-3}$ / $5.34 \times 10^{-4}$
Quebec, Canada	336	$4.84 \times 10^{-3}$	0.48 / 0.51	534 / 73	$1.79 \times 10^{-2}$ / $2.97 \times 10^{-3}$

<sup>1</sup>Calculated using the same mechanism as in section 4.3

<sup>2</sup>Measured in days since February 18, 2020

<sup>3</sup>Excluding simulations which do not predict second waves

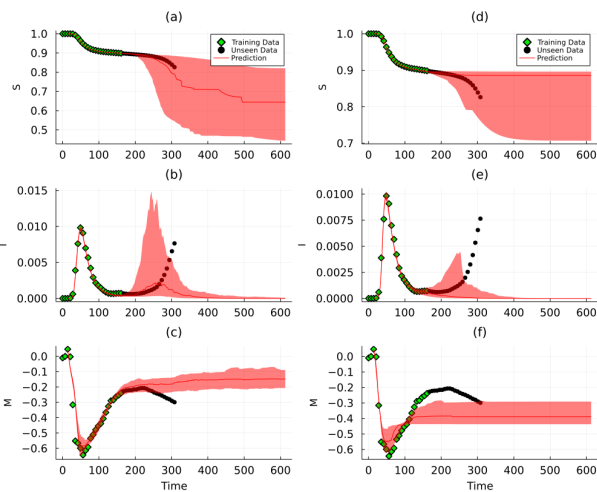
The model typically estimates that a 40-50% reduction in mobility is necessary to reduce  $R_0$  below 1. This is consistently more extreme than other studies have found (20-40%) [62], but not entirely implausible considering the interquartile range. That said, we cannot interpret any result for Belgium, California, or the UK where the interquartile range exceeds physically realistic bounds.

## Discussion

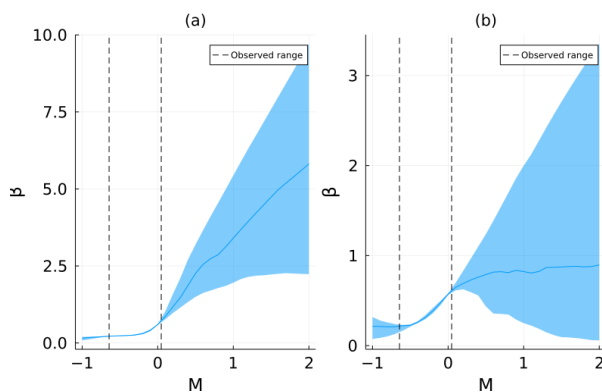
Our results show that socially and biologically informed machine learning models can perform qualitative prediction

tasks. When supplied with learning biases, the model routinely predicted second pandemic waves similar to those that occurred in most populations during the COVID-19 pandemic. The model seldom produced implausible predictions for mobility, and where it did, this tended to result from a failure to converge during training.

The most significant result is that the biased model predicts a second wave in every region except the Canadian province of British Columbia. The biased model predicted the second wave peak more consistently and closer to the actual time than the unbiased model without behavioural (mobility) feedback. The biased model also tended to predict a second wave that was much larger than the first wave, as occurred in most populations



**Fig. 2.** Predicted time series of all model states for New York state with learning biases (a-c) and without (d-f). Panels (a) and (b) show susceptible fraction, (b) and (d) show infected, and (c) and (f) show mobility. Green dots represent training data, while black dots represent unseen data. All predictions are generated using the median (solid line) and interquartile range (ribbon) of 100 independently-trained instances of the model.



**Fig. 3.** Predicted force of infection based on mobility level for New York state with learning biases (a) and without (b). Dotted lines indicate values of mobility seen by the model during training. Solid line shows the median prediction of 100 model instances and ribbon shows interquartile range.

during the COVID-19 pandemic, although the predicted second wave was often larger in magnitude than what occurred in reality.

This ability to predict second waves is valuable from a public health perspective, for mitigation of population health impacts. Though our model does not explicitly include government policy, it can influence behaviour, and knowing the likely trajectory of future cases under current policy can help decision-makers assess whether mandates should be tightened or loosened [36, 34]. In practice, our model could be used to simulate possible outcomes by using the trained  $\beta$  network, but changing  $M$  to a time signal representing total lifting of restrictions, gradual reopening, or continuing heavy restriction. Such a model may need to account in some way for the costs of each policy.

Mixed machine learning models need not supplant traditional models entirely, but they can be a valuable auxiliary. As our model shows, they need not be overly complex or

**Table 2.** Predicted  $R_0$  and required mobility reduction

Region	Predicted $R_0$	$M_{crit}$	Convergence <sup>1</sup>
Austria	2.2 ( $\pm 0.2$ )	-0.44 ( $\pm 0.20$ )	0.96
Belgium	2.6 ( $\pm 1.2$ )	-0.56 ( $\pm 0.54$ )*	0.76
Germany	2.6 ( $\pm 0.3$ )	-0.51 ( $\pm 0.21$ )	0.98
Italy	2.1 ( $\pm 0.5$ )	-0.49 ( $\pm 0.19$ )	0.99
Netherlands	2.1 ( $\pm 0.2$ )	-0.49 ( $\pm 0.17$ )	0.99
United Kingdom	1.9 ( $\pm 0.6$ )	-0.17 ( $\pm 7.99$ )*	0.79
California, USA	1.6 ( $\pm 0.9$ )	-0.44 ( $\pm 3.96$ )*	0.6
New York, USA	2.4 ( $\pm 0.2$ )	-0.52 ( $\pm 0.19$ )	0.92
Pennsylvania, USA	2.3 ( $\pm 0.7$ )	-0.56 ( $\pm 0.14$ )	0.94
Texas, USA	1.7 ( $\pm 0.8$ )	-0.57 ( $\pm 0.52$ )	0.82
British Columbia	1.6 ( $\pm 0.1$ )	-0.44 ( $\pm 0.22$ )	0.99
Ontario	2.4 ( $\pm 0.7$ )	-0.65 ( $\pm 0.21$ )	0.80
Quebec	2.3 ( $\pm 1.2$ )	-0.53 ( $\pm 0.30$ )	0.92

Parentheses show 95% credible intervals

<sup>1</sup>Not all neural networks showed a root. Interval calculated using only those that did.

\*Interval exceeds physically realistic values

computationally expensive. They can interpret large amounts of data, generalize well to a variety of different regions and, given appropriate learning biases, can be relied upon to make feasible predictions.

Epidemic models are often under-determined by data [63]. UDEs allow a new approach to this problem. Since neural networks are universal approximators [64], they can represent the full range of possible functions that could fit the available data. By training multiple iterations of a UDE model and analyzing their trajectories, we can see a range of feasible outcomes for the system with just one single model. For example, two UDE predictions can fit the data and biological constraints equally well, yet one may predict a massive second wave, while the other predicts a rapid return to normalcy. A third may produce several smaller waves with corresponding mobility changes. That said, it is important to assign sufficient weight to the learning biases to avoid discouraging such a range of behaviours in favour of a single, overfitted solution.

The ability of UDEs to examine a range of data-fitting functions could be further enhanced with sparse regression methods [51, 65, 66]. By applying sparse regression to our trained  $\beta$  model's output, one could derive a multitude of symbolic equations that could be used to mathematically model the system.

The results also support our hypothesis that learning biases are effective at accelerating training and assuring socially and biologically plausible solutions while achieving superior training performance. While some attributes can be learned passably well by the unbiased model given sufficient training time, the biased model still achieves better losses on these attributes by at least two orders of magnitude. Good performance on training data should not be taken too seriously since it may be a sign of overtraining. However, this does not appear to be the case in our model. The vast majority of the average training loss comes from a few highly divergent solutions. The improved performance by the biased model indicates reduced proclivity for such solutions.

The fact that a monotonic  $\beta$  is learned comparably well by both models indicates that both of them are instrumentally useful for the model to learn when satisfying the loss objective. This gives a good sanity check that these features are present in

the real system and assuming them in the model is reasonable. The upper and lower bounds on mobility, however, are not typically inferred by the model without explicit instruction. This is not unexpected since the observed data never nears the bounds. By fitting the data well, the model never needs to learn what happens at those bounds. However, including these boundaries as learning biases gives greater assurance that the model will not produce divergent or unstable solutions if, for example, it were used to predict what would happen in a scenario where those bounds were neared. Of course, it is preferable to ensure stability mathematically using structural biases, but this may not always be feasible. The other objective which the unbiased model tends not to learn is in the tendency for  $M$  to return to the baseline value of 0. This may be because, as people become accustomed to life with the virus, the "baseline," i.e. the average societal preference in the absence of disease, shifts downward.

It is interesting that the learning biases help generate greater variability in the out-of-sample time series predictions. This is likely because in the absence of any other objective, the model consistently converges to a single global optimum for data fitting. Since the model's extended prediction tends to remain within a small region of state space ( $M$  remaining negative,  $I$  relatively small), the greater potential variance is never realized. The model with learning biases, meanwhile, is relatively less concerned with fitting the data and hence has more freedom to explore the parameter space. The fact that the constraint losses are evaluated according to randomly generated sample points also confers greater variability to the results of the biased model.

The biased model also has greater variability in the upper quartile of its  $\beta$  response, but reduced variability in the lower quartile. This makes sense – the biased model has learned that for any  $M$  greater than those it has seen, the value of  $\beta$  must also be greater (and vice-versa for  $M$  less than what it has seen). The unbiased model, having no such information, cannot make an informed prediction, and so is equally likely to predict continued increase or an unrealistic decrease.

These variability trade-offs favour the biased model. Greater variability in time series prediction is valuable because (assuming the predictions are biologically feasible) it shows a greater variety of possible outcomes and assigns a degree of confidence associated to those outcomes. The reduced variability in predicting the transmissibility is also desirable because it derives from better understanding of the system.

We used a heavily simplified model of COVID-19. It is not intended to capture all details of the pandemic, nor is it meant to recommend specific health policies. We assumed the acquired immunity is permanent, which it may well not be [67]. We do not account for vaccines, which came into play around the end of 2020 [68]. Thus, the long-term predictions (i.e. beyond 300 days or so) should be taken only as evidence that the model does not produce wildly implausible behaviour, rather than a serious attempt to forecast cases too far in the future. The emergence of new variants, first reported in September 2020 [69] at the end of the second wave in many populations, mean that predictions for the tail end of 2020 are beyond the model's intended scope.

Even in the short term, the model is not intended to predict cases or to precisely estimate the virus's basic reproduction number. It is limited by our ability to consistently measure recovery rates and estimate under-reporting ratios, which almost certainly vary between regions and over time within regions. For simplicity, we also left out asymptomatic

transmission and seasonal changes in infectiousness, both of which hamper the model's short-term predictive ability compared to a more complex model [63]. None of these limitations changed our conclusions, since our goal was to show that UDEs and PIML can fit available data while making qualitatively correct out-of-sample predictions.

Future work could improve our the model by incorporating some of the aforementioned details of the pandemic. This could give insight into other behaviour-disease interactions like vaccine usage [70] or allow an examination of how these dynamics changed over the course of the pandemic. In section 4.1-4.2.1 we also provide some methodological changes that could further develop the UDE/PIML themes, particularly regarding how to use learning biases effectively.

Probably the biggest opportunity for future work is to apply this type of data-driven differential equation model to other systems. Other infectious diseases, particularly those for which vaccines are available, are also coupled behavior-disease systems [70, 66] and so could be amenable to this type of model. Beyond epidemic modelling, climate systems are also known to have important behavioural components [71, 72]. Ultimately, one of the greatest advantages of UDEs is that, as per their name, they can theoretically be applied to any dynamical system [51]. It is only a matter of testing them to see if they provide valuable insight which.

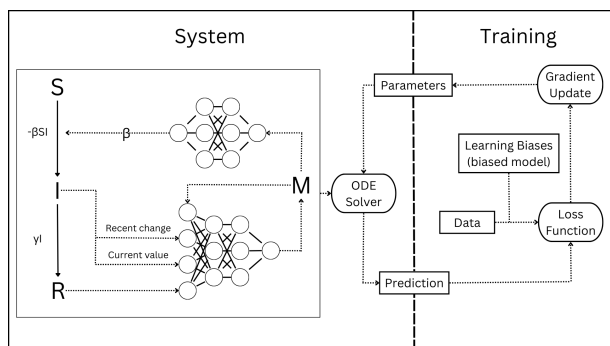
Availability of social and epidemiological data for endemics and future pandemics will likely continue to increase. At the same time, socio-economic factors will continue being complex, and regional and temporal variability will persist. Universal differential equations, when supplied with appropriate learning biases, could be a valuable tool for modelling such systems. Alongside traditional models, they could be used to quickly gain perspective on the state of outbreaks across the world without having to develop specialized models for each region.

## Materials and Methods

Our model is a universal delay differential equation (UDDE) based on the standard SIR model:

$$\begin{aligned}\frac{d}{dt}S(t) &= -\beta(M(t - \tau_1))S(t)I(t) \\ \frac{d}{dt}I(t) &= \beta(M(t - \tau_1))S(t)I(t) - \gamma I(t) \\ \frac{d}{dt}M(t) &= e^{-\delta t} f(I(t - \tau_1), \Delta I(t - \tau_2), M(t), R(t))\end{aligned}\quad (1)$$

where  $S$ ,  $I$  and  $R$  represent the susceptible, infected and recovered proportions of the population respectively ( $R$  can be recovered as  $1 - S - I$ ), and  $M$  represents the percent difference in mobility compared to the baseline.  $\beta(M)$  represents the transmission rate as it depends on mobility, and  $f(I, \Delta I, M, R)$  represents the dynamics governing social/behavioural (mobility) response to the infectious disease [35]. Both  $\beta$  and  $f$  were learned by the algorithm.  $\Delta I(t)$  represents the change in  $I$  between the current time  $t$  and a previous time  $t - \tau_2$ . The  $e^{-\delta t}$  factor accounts for several factors that reduce the population response to the virus over time, including: pandemic fatigue, the development of medical interventions that make the infection less fatal (such as 'proning'), substituting less disruptive interventions (such as masking) for mobility reductions, and (for longer-term predictions than we study in this model) the evolution to milder



**Fig. 4.** Schematic of our model showing the relevant differential equations, neural networks, and training procedure. The learning biases are present only in the biased model. Otherwise, the biased and unbiased models have the same structure.

virulence over time. Section 4 Figure 4 shows a schematic of the model.

The model parameters are  $\gamma$ , the per-capita recovery rate,  $\tau_M$ , the delay between a change in  $M$  and the corresponding change in prevalence, and  $\tau_R$ , the reverse delay: the time between a change in prevalence and corresponding behavioural response [35]. The values we used are  $\gamma = 0.25\text{day}^{-1}$  [73],  $\tau_M = 10\text{days}$ , and  $\tau_I = 14\text{days}$  [74, 62].

This model inherits several structural biases from the standard SIR model template. First,  $S = 0$  and  $I = 0$  are both invariant, preventing any unbiological negative values for these variables. Second, it retains the conservation relation  $S + I + R = 1$ . Thus, regardless of the functions fit by the neural network,  $S(t)$  and  $I(t)$  are guaranteed to be plausible.

### Neural Networks

The influence of mobility (i.e. contact rate) on the transmission rate is represented by neural networks that are used to represent  $\beta(M)$  and  $f(I, \Delta I, M, R)$ . These networks each have linear output layers with one neuron and 2 hidden layers with 3 neurons per hidden layer and Gaussian Error Linear Unit (GELU) activation functions. This gives the  $\beta$  network 22 parameters and the  $f$  network 31, for a total of 54 once  $\delta$  is included. Different network structure (size, activation function) could have improved performance. The hyperparameter space is probably too large to globally optimize the model, but there may be improvements we missed in our hyperparameter tuning.

### Training Methodology

The baseline (unbiased) model, which received no social or biological feedback, was trained only to fit the data (details in section 4.3). The model's prediction is generated by solving the delay differential equation system to get its prediction for each state at each time step. This prediction is then fed into a scaled mean-squared error loss function:

$$L(\Theta) = \sum_{i=1}^n \sum_{j=1}^m \left( \frac{y_{ij} - \bar{y}_{ij}}{m(\bar{y}_{i \max} - \bar{y}_{i \min})} \right)^2 \quad (2)$$

Here,  $n$  is the dimension of the system,  $m$  is the number of data points,  $y_{ij}$  is the true value of the  $i$ th variable's  $j$ th data point, and  $\bar{y}_{ij}$  is the prediction for  $i$ th variable's  $j$ th data point.  $k$  is the size of the parameter vector  $\Theta$ , and  $\Theta_l$  is the  $l$ th entry in  $\Theta$ . Scaling the loss function in this way helps ensure all variables are given equal importance despite having different ranges [75].

Both biased and unbiased models for all regions were trained on the first 160 days, giving  $m = 22$  data points after weekly averaging (see 4.3). This time period fully encompasses the first wave for all populations studied, but does not include the beginning of the second wave.

### Learning Biases

The socially and biologically informed model was trained to minimize the same accuracy loss objective as well as 8 other objectives, each encoding a social or biological assumption. These BIML loss functions are deliberately constructed so as to give 0 loss to any functions which satisfy the relevant assumptions. This allows the model complete freedom to explore the range of biologically feasible functions.

At each training iteration, the model generates 100 random points in the region  $0 \leq I \leq 1, I - 1 \leq \Delta I \leq I, -100 \leq M \leq 100$ . The BIML loss functions are evaluated at each of these points. The total loss at each iteration is then a weighted sum of the BIML losses, plus the accuracy loss. We tried dynamically updating the weights for each loss function as in [25], but this did not significantly improve results. The BIML assumptions and corresponding loss functions are displayed in Section 4.2.1 Table 3.

We only tested a few values for the learning bias weight. The optimal value for achieving tolerable performance on training data while assuring qualitatively realistic long-term predictions may be higher or lower. It may instead involve weights that differ between loss functions or which evolve over the training process.

Model parameters were randomly initialized. Parameter sets that gave initial errors of more than  $10^4$  were re-initialized. We used the Adam optimizer to optimize the model parameters. For the basic model, we simply use the gradient of the accuracy loss with respect to the parameters. For the biologically-informed model, we use the same weighted sum for the total gradient as for the total loss. More details on the software implementation of this algorithm can be found in the appendix.

We found that training the model on the entire training set at once caused it to become stuck in a local optimum where  $I$  never increased. Thus, we trained the models in stages to achieve a better fit more quickly. The model trained on the first quarter of the data in the first stage, the first half in the second, and the entire training set in the third.

Repeating our model with more computing time and power could be informative. Although we were able run the model with enough iterations to ensure all models converged to a good degree, some certainly converged better than others. The mobility data was a particular challenge, with fairly sharp downturns and upturns occasionally not always fully captured. This could be assisted by using collocation-based training to speed up the process [76].

### Data

#### Case Data

Daily case notification data was taken from the Johns Hopkins CSSE dataset [77]. We derived daily susceptible and infected proportions using the following system:

$$\begin{aligned} I_T(t_n) &= I_T(t_{n-1}) - \gamma I_T(t_{n-1}) + 5C(t_n) \\ S_T(t_n) &= S_T(t_{n-1}) - 5C(t_n) \end{aligned} \quad (3)$$

where  $S_T(t_n)$  and  $I_T(t_n)$  are the total number of susceptible and infected individuals (not proportions) on day  $t_n$  and  $C(t_n)$

**Table 3.** BINN loss functions

Biological assumption	Loss function
1. $\beta(M) \geq 0$ for all $M$	$\text{relu}(-\beta(M_i))$
2. Higher values of $M$ correspond to higher values of $\beta(M)$	$\text{relu}(\beta(M_i)) - (\beta(M_i + \epsilon))$
3. $M$ cannot go below a minimum value, i.e. a reduction of 100% from baseline	$\text{relu}(-f(M_{\min}, I_i, \Delta I_i, R))$
4. $M$ cannot exceed a maximum value, set <sup>1</sup> at $M_{\max} = 2(M_{\text{baseline}} - M_{\min})$	$\text{relu}(f(M_{\max}, I_i, \Delta I_i, R))$
5. $f$ is monotonically decreasing in $I$	$\text{relu}(f(M_i, I_i + \epsilon, \Delta I_i, R) - f(M_i, I_i, \Delta I_i, R))$
6. $f$ is monotonically decreasing in $\Delta I$	$\text{relu}(f(M_i, I_i, \Delta I_i + \epsilon, R) - f(M_i, I_i, \Delta I_i, R))$
7. In the absence of infection, $M$ tends towards baseline	$\text{relu}[f(M, I, \Delta I, R)(M_i - M_{\text{baseline}})]$
8. $M$ tends toward baseline more strongly when $R$ is higher	$\text{relu}[ f(M, I, \Delta I, R) -  f(M, I, \Delta I, R + \epsilon) ]$

<sup>1</sup>This upper bound is deliberately (if arbitrarily) set much higher than the maximum of the data to stop solutions diverging to infinity without excessively constraining the model

are the number of new cases on day  $t_n$ . The parameter  $\gamma$  is the same used in the model as discussed above. Finally, we divided  $I$  and  $S$  by the total population to get a proportion at each time step to ensure different regions are comparable.

#### Mobility Data

Daily mobility data ( $M$ ) comes from the Google Community Mobility Report [78]. We mean-normalized this data to give it a comparable range to  $S$  and  $I$ . We chose the Retail and Recreation category of mobility data as it corresponded most to what we were trying to measure: voluntary activities in indoor settings that place people at risk of becoming infected. It is also strongly correlated with infection, so it is reasonable to expect it can be a good predictor of cases. Repeating the simulations with workplace mobility would be a good test of the model's validity. Other mobility measures, such as parks, would be difficult to use due to their weaker correlation with infections [17].

Once we had data for each day, we sub-sampled it by taking a weekly moving average to reduce irregularities from weekends, holidays, and days where data was not available. We set the initial condition as the first data point for which  $I \neq 0$ .

The regions analyzed were chosen to represent a variety of relatively populous locations across Western Europe, the US, and Canada. We chose to focus particularly on regions where populations are decently concentrated and which had notable second waves.

#### Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions.

#### Supplementary Material

Supplementary material is available at PNAS Nexus online.

#### Funding

This work was supported by an NSERC Discovery Grant to C.T.B., and an NSERC Undergraduate Student Research Assistant Award to B.K.

#### Author contributions statement

C.T.B. conceived the project, provided guidance and direction, conducted and shared preliminary research, and reviewed the

manuscript. B.K. designed the specific details of the model, coded and ran the simulations, and analyzed the results. Both authors contributed to writing the manuscript.

#### Data availability

The code used to run and analyze the model is available at <https://github.com/bkuwahara/predicting-2ndwaves-udes>. All empirical data are publicly available (see References).

#### References

1. Alessandro Vespignani, Huaiyu Tian, Christopher Dye, James O Lloyd-Smith, Rosalind M Eggo, Munik Shrestha, Samuel V Scarpino, Bernardo Gutierrez, Moritz UG Kraemer, Joseph Wu, et al. Modelling covid-19. *Nature Reviews Physics*, 2(6):279–281, 2020.
2. Julia R Gog. How you can help with covid-19 modelling. *Nature Reviews Physics*, 2(6):274–275, 2020.
3. C Jessica E Metcalf, Dylan H Morris, and Sang Woo Park. Mathematical models to guide pandemic response. *Science*, 369(6502):368–369, 2020.
4. Meagan C Fitzpatrick, Chris T Bauch, Jeffrey P Townsend, and Alison P Galvani. Modelling microbial infection to address global health challenges. *Nature microbiology*, 4(10):1612–1619, 2019.
5. Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.
6. Karen KL Hwang, Christina J Edholm, Omar Saucedo, Linda JS Allen, and Nika Shakiba. A hybrid epidemic model to explore stochasticity in covid-19 dynamics. *Bulletin of Mathematical Biology*, 84(9):91, 2022.
7. Estee Y Cramer, Evan L Ray, Velma K Lopez, Johannes Bracher, Andrea Brennen, Alvaro J Castro Rivadeneira, Aaron Gerding, Tilmann Gneiting, Katie H House, Yuxin Huang, et al. Evaluation of individual and ensemble probabilistic forecasts of covid-19 mortality in the united states. *Proceedings of the National Academy of Sciences*, 119(15):e21113561119, 2022.
8. Kathryn R. Fair, Vadim A. Karatayev, Madhur Anand, and Chris T. Bauch. Estimating covid-19 cases and deaths prevented by non-pharmaceutical interventions, and the impact of individual actions: A retrospective model-based analysis. *Epidemics*, 39:100557, 2022.
9. Chad R Wells, Jeffrey P Townsend, Abhishek Pandey, Seyed M Moghadas, Gary Krieger, Burton Singer, Robert H McDonald, Meagan C Fitzpatrick, and Alison P Galvani.



- Optimal covid-19 quarantine and testing strategies. *Nature communications*, 12(1):356, 2021.
10. Ashleigh R. Tuite, David N. Fisman, and Amy L. Greer. Mathematical modelling of covid-19 transmission and mitigation strategies in the population of ontario, canada. *Canadian Medical Association Journal*, 192(19):E497–E505, 2020.
  11. Jude Bayham and Eli P Fenichel. Impact of school closures for covid-19 on the us health-care workforce and net mortality: a modelling study. *The Lancet Public Health*, 5(5):e271–e278, 2020.
  12. Peter C Jentsch, Madhur Anand, and Chris T Bauch. Prioritising covid-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study. *The Lancet Infectious Diseases*, 21(8):1097–1106, 2021.
  13. Seyed M Moghadas, Thomas N Vilches, Kevin Zhang, Chad R Wells, Affan Shoukat, Burton H Singer, Lauren Ancel Meyers, Kathleen M Neuzil, Joanne M Langley, Meagan C Fitzpatrick, et al. The impact of vaccination on coronavirus disease 2019 (covid-19) outbreaks in the united states. *Clinical Infectious Diseases*, 73(12):2257–2264, 2021.
  14. Caroline E Wagner, Chadi M Saad-Roy, and Bryan T Grenfell. Modelling vaccination strategies for covid-19. *Nature Reviews Immunology*, 22(3):139–141, 2022.
  15. Dongshan Zhu, Xikun Han, and Karla Santo. Social distancing in latin america during the covid-19 pandemic: an analysis using the stringency index and google community mobility reports. *Journal of Travel Medicine*, 27(8), 2020.
  16. Antonio Paez. Using google community mobility reports to investigate the incidence of covid-19 in the united states. *Findings*, 5 2020.
  17. M. Sulyok and M. Walker. Community movement and covid-19: a global study using google’s community mobility reports. *Epidemiology and Infection*, 148:e284, 2020.
  18. Vadim A. Karatayev, Madhur Anand, and Chris T. Bauch. Local lockdowns outperform global lockdown on the far side of the covid-19 epidemic curve. *Proceedings of the National Academy of Sciences*, 117(39):24575–24580, 2020.
  19. Yehuda Pollak, Haym Dayan, Rachel Shoham, and Itai Berger. Predictors of adherence to public health instructions during the covid-19 pandemic. *medRxiv*, 2020.
  20. Abel Brodeur, Andrew E. Clark, Sarah Fleche, and Nattavudh Powdthavee. Covid-19, lockdowns and well-being: Evidence from google trends. *Journal of Public Economics*, 193:104346, 2021.
  21. Isaac Yen-Hao, Prima Alam, Heidi. J Larson, and Leesa Lin. Social consequences of mass quarantine during epidemics: a systematic review with implications for the covid-19 response. *Journal of Travel Medicine*, 27(7), 2020.
  22. Zhen Wang, Michael A Andrews, Zhi-Xi Wu, Lin Wang, and Chris T Bauch. Coupled disease–behavior dynamics on complex networks: A review. *Physics of life reviews*, 15:1–29, 2015.
  23. Chris Bauch, Alberto d’Onofrio, and Piero Manfredi. Behavioral epidemiology of infectious diseases: an overview. *Modeling the interplay between human behavior and the spread of infectious diseases*, pages 1–19, 2013.
  24. Sebastian Funk, Marcel Salathé, and Vincent AA Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7(50):1247–1256, 2010.
  25. Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why pinns fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022.
  26. Timothy C. Reluga. Game theory of social distancing in response to an epidemic. *PLOS Computational Biology*, 6(5):1–9, 05 2010.
  27. Alina Glaubitz and Feng Fu. Oscillatory dynamics in the dilemma of social distancing. *Proceedings of the Royal Society A*, 476(2243):20200686, 2020.
  28. Chris T Bauch and David JD Earn. Vaccination and the theory of games. *Proceedings of the National Academy of Sciences*, 101(36):13391–13394, 2004.
  29. Alberto d’Onofrio, Piero Manfredi, and Ernesto Salinelli. Vaccinating behaviour, information, and the dynamics of sir vaccine preventable diseases. *Theoretical population biology*, 71(3):301–317, 2007.
  30. Iain R Moyles, Jane M Heffernan, and Jude D Kong. Cost and social distancing dynamics in a mathematical model of covid-19 with application to ontario, canada. *Royal Society open science*, 8(2):201770, 2021.
  31. KM Ariful Kabir and Jun Tanimoto. Evolutionary game theory modelling to represent the behavioural dynamics of economic shutdowns and shield immunity in the covid-19 pandemic. *Royal Society open science*, 7(9):201095, 2020.
  32. Joshua M Epstein, Erez Hatna, and Jennifer Crodelle. Triple contagion: a two-fears epidemic model. *Journal of the Royal Society Interface*, 18(181):20210186, 2021.
  33. Bruno Buonomo, Rossella Della Marca, Alberto d’Onofrio, and Maria Groppi. A behavioural modelling approach to assess the impact of covid-19 vaccine hesitancy. *Journal of Theoretical Biology*, 534:110973, 2022.
  34. Sansao A. Pedro, Frank T. Ndjomatchoua, Peter Jentsch, Jean M. Tchuenche, Madhur Anand, and Chris T. Bauch. Conditions for a second wave of covid-19 due to interactions between disease dynamics and social processes. *Frontiers in Physics*, 8, 2020.
  35. Matthew D. Johnston and Bruce Pell. A dynamical framework for modeling fear of infection and frustration with social distancing in covid-19 spread, 2020.
  36. Ashleigh R Tuite and Amy L Greer. Risk for covid-19 resurgence related to duration and effectiveness of physical distancing in ontario, canada. *Annals of Internal Medicine*, 173(8):675–678, 2020. PMID: 32459528.
  37. Iman Rahimi, Fang Chen, and Amir H Gandomi. A review on covid-19 forecasting models. *Neural Computing and Applications*, 2021.
  38. Maximilian Nickel, Levent Sagun, Mark Ibrahim, Matt Le, and Timothee Lacroix. Neural relational autoregression for high-resolution covid-19 forecasting. *Meta AI ML Applications*, 2020.
  39. Raj Dandekar, Chris Rackauckas, and George Barastathis. A machine learning-aided global diagnostic and comparative tool to assess effect of quarantine control in covid-19 spread. *Patterns*, 1(9), 2020.
  40. Gergo Pinter, Imre Felde, Amir Mosavi, Pedram Ghamisi, and Richard Gloaguen. Covid-19 pandemic prediction for hungary; a hybrid machine learning approach. *Mathematics*, 8(6), 2020.
  41. Sina F. Ardabili, Amir Mosavi, Pedram Ghamisi, Filip Ferdinand, Annamaria R. Varkonyi-Koczy, Uwe Reuter, Timon Rabczuk, and Peter M. Atkinson. Covid-19 outbreak prediction with machine learning. *Algorithms*, 13(10), 2020.

42. George E Karniadakis, Ioannidis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
43. Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 11 2018.
44. Sharmila Karumuri, Rohit Tripathy, Ilias Bilionis, and Jitesh Panchal. Simulator-free solution of high-dimensional stochastic elliptic partial differential equations using deep neural networks. *Journal of Computational Physics*, 404:109120, 2020.
45. Dongkun Zhang, Ling Guo, and George Em Karniadakis. Learning in modal space: Solving time-dependent stochastic pdes using physics-informed neural networks. *SIAM Journal on Scientific Computing*, 42(2):A639–A665, 2020.
46. Georgios Kissas, Yibo Yang, Eileen Hwuang, Walter R. Witschey, John A. Detre, and Paris Perdikaris. Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4d flow mri data using physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 358:112623, 2020.
47. Casey S. Greene and James C. Costello. Biologically informed neural networks predict drug responses. *Cancer Cell*, 38(5):613–615, 2020.
48. Haitham A Elmarakeby, Justin Hwang, Rand Arafah, Jett Crowdis, Sydney Gang, David Liu, Saud H Al Dubayan, Keyan Salari, Steven Kregel, Camden Richter, Taylor E Arnoff, Jihya Park, William C Hahn, and Eliezer M Van Allen. Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598:348–352, 2021.
49. Andrew Beers, Ken Chang, James M. Brown, Emmett Sartor, C. P. Mammen, Elizabeth R. Gerstner, Bruce R. Rosen, and Jayashree Kalpathy-Cramer. Sequential 3d unets for biologically-informed brain tumor segmentation. *CoRR*, abs/1709.02967, 2017.
50. John H. Lagergren, John T. Nardini, Ruth E. Baker, Matthew J. Simpson, and Kevin B. Flores. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *PLOS Computational Biology*, 16(12):1–29, 12 2020.
51. Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, and Ali Jasim Ramadhan. Universal differential equations for scientific machine learning. *CoRR*, abs/2001.04385, 2020.
52. Momchil Minkov, Ian A. D. Williamson, Lucio C. Andreani, Dario Gerace, Beicheng Lou, Alex Y. Song, Tyler W. Hughes, and Shanhui Fan. Inverse design of photonic crystals through automatic differentiation. *ACS Photonics*, 7(7):1729–1741, 2020.
53. Raj Dandekar and George Barbastathis. Neural network aided quarantine control model estimation of covid spread in wuhan, china, 2020.
54. Junling Ma. Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*, 5:129–141, 2020.
55. Robert M. May. Uses and abuses of mathematics in biology. *Science*, 303(5659):790–793, 2004.
56. Chris T Bauch. Estimating the COVID-19 R number: a bargain with the devil? *The Lancet Infectious Diseases*, 21(2):151–153, February 2021. Publisher: Elsevier.
57. Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*, 27(2), 2020.
58. Nithya C Achaiah, Sindhu B Subbarajasetty, and Rajesh M Shetty. R0 and re of covid-19: Can we predict when the pandemic outbreak will be contained? *Indian Journal of Critical Care Medicine*, 2020.
59. Vishal Srivastava, Smriti Srivastava, Gopal Chaudhary, and Fadi Al-Turjman. A systematic approach for covid-19 predictions and parameter estimation. *Personal and Ubiquitous Computing*, 2020.
60. Weston C. Roda, Marie B. Varughese, Donglin Han, and Michael Y. Li. Why is it difficult to accurately predict the covid-19 epidemic? *Infectious Disease Modelling*, 5:271–281, 2020.
61. Kevin Linka, Mathias Peirlinck, and Ellen Kuhl. The reproduction number of covid-19 and its correlation with public health interventions, 2020.
62. Robert B. Noland. Mobility and the effective reproduction rate of covid-19. *Journal of Transport & Health*, 20:101016, 2021.
63. Hyun Mo Yang, Luis Pedro Lombardi Junior, and Ariana Campos Yang. Are the sir and seir models suitable to estimate the basic reproduction number for the covid-19 epidemic? *medRxiv*, 2020.
64. Sho Sonoda and Noboru Murata. Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, 43(2):233–268, 2017.
65. Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
66. Jonathan Horrocks and Chris T Bauch. Algorithmic discovery of dynamic models from infectious disease data. *Scientific Reports*, 10(1):7061, 2020.
67. Chris Baraniuk. How long does covid-19 immunity last? *BMJ*, 373, 2021.
68. AJMC Staff. A timeline of covid-19 vaccine developments in 2021. @ONLINE, June 2021.
69. World Health Organization. Tracking sars-cov-2 variants @ONLINE, November 2021.
70. Chris T. Bauch and David J. D. Earn. Vaccination and the theory of games. *Proceedings of the National Academy of Sciences*, 101(36):13391–13394, 2004.
71. Alison P. Galvani, Chris T. Bauch, Madhur Anand, Burton H. Singer, and Simon A. Levin. Human–environment interactions in population and ecosystem health. *Proceedings of the National Academy of Sciences*, 113(51):14502–14506, 2016.
72. Thomas M Bury, Chris T Bauch, and Madhur Anand. Charting pathways to climate change mitigation in a coupled socio-climate model. *PLoS computational biology*, 15(6):e1007000, 2019.
73. Hiroshi Nishiura, Natalie M Limton, and Andrei R Akhmetzhanov. Serial interval of novel coronavirus (covid-19) infections. *International Journal of Infectious Diseases*, 93:284–286, 2020.
74. Hamada S Badr, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M Squire, and Lauren M

- Gardner. Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(11):1247–1254, 2020.
75. Suyong Kim, Weiqi Ji, Sili Deng, Yingbo Ma, and Christopher Rackauckas. Stiff neural ordinary differential equations. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(9):093122, 2021.
76. Elisabeth Roesch, Christopher Rackauckas, and Michael P. H. Stumpf. Collocation based training of neural ordinary differential equations. *Statistical Applications in Genetics and Molecular Biology*, 20(2):37–49, 2021.
77. [Dataset]\* Johns Hopkins University. Covid-19 data repository by the center for systems science and engineering (csse) at johns hopkins university.
78. [Dataset]\* Google. Google community mobility report.