
AutoODE: Bridging Physics-based and Data-driven modeling for COVID-19 Forecasting

Rui Wang
UC San Diego
ruw020@ucsd.edu

Danielle Maddix
Amazon Research
dmmaddix@amazon.com

Christos Faloutsos
Amazon and CMU
faloutso@amazon.com

Yuyang Wang
Amazon Research
yuyawang@amazon.com

Rose Yu
UC San Diego
roseyu@eng.ucsd.edu

Abstract

As COVID-19 continues to spread, accurately forecasting the number of newly infected, removed and death cases has become a crucial task in public health. While mechanics compartment models are widely-used in epidemic modeling, data-driven models are emerging for disease forecasting. In this work, we investigate these two types of methods for COVID-19 forecasting. Through a comprehensive study, we find that data-driven models outperform physics-based models on the number of death cases prediction. Meanwhile, physics-based models have superior performances in predicting the number of infected and removed cases. In addition, we present an hybrid approach, AutoODE, that obtains a 57.4% reduction in mean absolute errors of the 7-day ahead COVID-19 trajectories prediction compared with the best deep learning competitor.

1 Introduction

Mechanistic models are commonly used in epidemiology to model the spreading of diseases because of their robustness and interpretability. However, physics-based models can be limited by their strong assumptions, such as a constant transmission rate, and may be not flexible enough to model the complex COVID-19 dynamics. For example, the SIR and SEIR compartmental models use coupled ODEs to model epidemic spreading, where the population is assigned to different states or compartments (Anderson et al., 1992). Limitations of applying the SEIR equations directly to COVID-19 are that it does not consider the unreported cases, and the effect of the exposed infecting susceptibles during the incubation period.

On the other hand, when f_θ in Eqn. (3.1) is unknown or too complex for ODEs to model, data-driven and model-free deep learning (DL) methods are preferable. These methods can leverage the abundant data and perform fast inference. Recently, DL models have seen great success in forecasting dynamical systems to extract information across multiple related time series. For instance, Chen et al. (2018); Wang et al. (2020b); Ayed et al. (2019) developed DL models based on differential equations for spatio-temporal dynamics forecasting. Raissi and Karniadakis (2018); Al-Arabi et al. (2018); Sirignano and Spiliopoulos (2018) tend to directly solve differential equations with DL given coordinates and time as input. It is known that DL methods adapt poorly to domain shifts, and cannot perform well with insufficient training data (Kouw and Loog, 2018; Amodei et al., 2019).

In this work, we perform a comprehensive benchmark study by comparing physics-based and deep learning methods. We focus on the task of forecasting the 7-day ahead COVID-19 trajectories.

We propose AutoODE, a hybrid method that estimates the unknown parameters θ in Eqn. (3.1) with auto-differentiation, and models the transmission rate as piecewise-constant to reflect the effect of various public health interventions. We also model the transmission among the US states with a low-rank approximated matrix learned from the data. As shown in Figure 1, the prediction

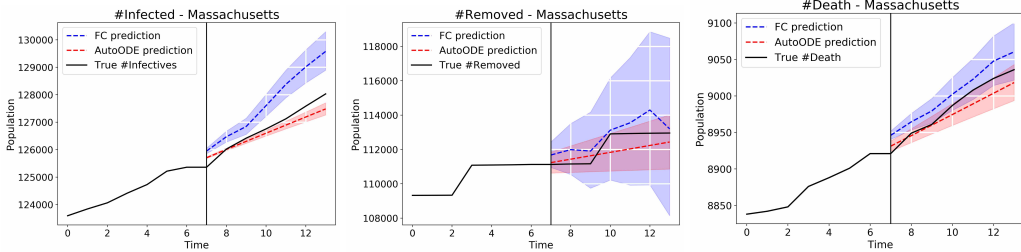


Figure 1: Proposed AutoODE wins: I , R and D predictions for week 08=23 – 08=29 in Massachusetts by our proposed AutoODE model and the best performing DL model FC.

by our AutoODE model is closer to the target 7-day ahead COVID predictions of I , R and D in Massachusetts than that from the best DL model (FC). In particular, AutoODE obtains a 57.4% reduction in mean absolute errors of 7-day ahead trajectories prediction compared with the best DL competitor.

2 Related Work

Deep Sequence Models. Deep sequence models, such as sequence to sequence (Seq2Seq) and Transformer models, have an encoder-decoder structure that can directly map input sequences to output sequences with different lengths (Vaswani et al., 2017; Wu et al., 2020; Li et al., 2020; Flunkert et al., 2017; Rangapuram et al., 2018). Fully connected (FC) neural networks can also be used autoregressively to produce multiple time-step forecasts (Benidis et al., 2020; Lim and Zohren, 2020). Another class of methods is based on the assumption that the data is governed by an ODE system, such as NeuralODE (Chen et al., 2018), which can generate predicted sequences of any length, since the ODE solver is able to generate continuous predictions. When the data is spatially correlated, such as traffic flow and COVID-19 trajectories among different states, deep graph models, such as graph convolution networks (GCN) and graph attention networks (GAN) (Velickovic et al., 2017), may be superior to deep sequence models since they are able to capture spatial structure.

Epidemic Forecasting. Compartmental models based ODEs have been widely used for modeling and forecasting epidemics. Chen et al. (2020) proposes a time-dependent SIR model that uses ridge regression to predict the transmission and recovery rates over time. A potential limitation with this method is that it does not consider the incubation period and unreported cases. Pei and Shaman (2020) modified the compartments in the SEIR model into the subpopulation commuting among different places, and estimated the model parameters using iterated filtering methods. Keskinocak et al. (2020) investigated the impact of social distancing on COVID-19 spread in Georgia. Wang et al. (2020a) proposes a population-level survival-convolution method to model the number of infectious people as a convolution of newly infected cases and the proportion of individuals remaining infectious over time after being infected. Zou et al. (2020) proposes the SuEIR model that incorporates the unreported cases, and the effect of the exposed group on susceptibles into the model. Davis et al. (2020) shows the importance of simultaneously modeling the transmission rate among the fifty U.S. states since transmission between states is significant.

3 Proposed method: AutoODE

We present our proposed AutoODE model that given an ODE in Eqn. (3.1) learns the unknown parameters with automatic differentiation using gradient-based methods. Unlike with neural networks, AutoODE is data-efficient, and the model only needs to be fit on the days before the prediction week. We apply this physics-based method to COVID-19 forecasting, using the ODEs in Eqn. (3.1) improved upon from the SuEIR model, where we estimate the unknown parameters β , γ , δ , and ρ , which denote the transmission, incubation, discovery, and recovery rates, respectively. The total population $N_i = S_i + E_i + U_i + I_i + R_i$ is assumed to be constant for each U.S. states i .

Low Rank Approximation to the Transmission Matrix: A_{ij} We introduce a transmission matrix A to model the transmission rate among the 50 U.S. states. Each entry of A is the element-wise

product of the sparse U.S. states adjacency matrix M and the correlation matrix C that is learned from data, that is, $A = C \cdot M \in \mathbb{R}^{n \times n}$. We omit the transmission between the states that are not adjacent to each other to avoid overfitting. To reduce the number of parameters and improve the computational efficiency to $O(kn)$, we use a low rank approximation to generate the correlation matrix $C = B^T D$, where $B; D \in \mathbb{R}^{k \times n}$ for $k \ll n$.

Piecewise Linear Transmission Rate: $\beta_i(t)$ Most compartmental models assume the transmission rate β_i is constant, which does not hold for COVID-19. The transmission rate of COVID-19 changes over time due to government regulations, such as school closures and social distancing. Even though we do short-term forecasting (7 days ahead), it is possible that the transmission rate may change during the training period. Instead of a constant approximation to β_i , we use a piecewise linear function over time $\beta_i(t)$, and set the breakpoints, slopes and biases as trainable parameters.

Death Rate Modeling: $r_i(t)$ Figure 3 in Appendix C shows the trajectories of the number of accumulated removed and death cases in four different states. We can see a relationship between the numbers of accumulated removed and death cases can be close to linear, exponential or concave. Since we do short-term forecasting, the death rate $r_i(t)$ can be assumed as a linear function $a_i t + b_i$ to cover both the convex and concave functions, where a_i and b_i are set as learnable parameters.

$$\begin{aligned}
 \frac{dS_i}{dt} &= \frac{\sum_{j=1}^n [\beta_i(t) A_{ij} (I_j + E_j) S_j]}{N_i}; \\
 \frac{dE_i}{dt} &= \frac{\sum_{j=1}^n [\beta_i(t) A_{ij} (I_j + E_j) S_j]}{N_i} - \beta_i E_i; \\
 \frac{dU_i}{dt} &= (1 - \beta_i) \beta_i E_i; \\
 \frac{dI_i}{dt} &= \beta_i E_i - \beta_i I_i; \\
 \frac{dR_i}{dt} &= \beta_i I_i; \\
 \frac{dD_i}{dt} &= r_i(t) \frac{dR_i}{dt}.
 \end{aligned} \tag{3.1}$$

Numerical Integration To solve the coupled ordinary differential equations, we use the 4-th order Runge-Kutta Method (RK4) (J.C. Butcher, 1996). In the Neural ODE method (Chen et al., 2018), the authors use the adjoint method to have the neural networks bypass the numerical solver, and be applicable to higher dimensional problems. In our case, since our method uses low dimension ordinary differential equations, RK4 is sufficient to generate accurate predictions. We directly implement RK4 in Pytorch, and allow backpropagation through it with a fixed time-step Δt .

Weighted Loss Function We set the unknown parameters in Eqn. (3.1) as trainable, and apply a gradient-based optimizer to minimize the following weighted loss function:

$$L(\mathbf{A}; \boldsymbol{\beta}; \boldsymbol{\sigma}; \boldsymbol{\mu}; \boldsymbol{\gamma}; \mathbf{r}) = \frac{1}{T} \sum_{t=1}^T w(t) [l_1(\hat{I}_t; I_t) + \lambda_1 l_2(\hat{R}_t; R_t) + \lambda_2 l_3(\hat{D}_t; D_t)];$$

with weights $\lambda_1; \lambda_2$ and loss function $l(\cdot; \cdot)$ to find the optimal parameters. We utilize these weights to balance the loss of the three states due to scaling differences, and also reweigh the loss at different time steps. We give larger weights to more recent data points by setting $w(t) = \frac{1}{T} \frac{1}{t}$. The constants, $\lambda_1; \lambda_2$ and T are tuned on the validation set. We set $l(\cdot; \cdot)$ to be the quantile loss (Wen et al., 2018) for both AutoODE and the DL models. Since I, R and D are the only known states in the real-world COVID-19 datasets, we show that our AutoODE model can correctly learn all the parameters and the trajectories of all six states based only on the loss I, R and D on synthetic data in Appendix B.

4 Experiments

Dataset We use the COVID-19 data from Apr 14 to Sept 12 provided by Johns Hopkins University Dong et al. (2020). It contains the number of cumulative number infected (I), recovered (R) and

death (D) cases. Figure 4 in Appendix C shows the rolling averages and standard deviation intervals of daily increase time series in New York, Pennsylvania, Maryland and Virginia.

Experimental Setup We investigate the following six DL models on forecasting COVID-19 trajectories, sequence to sequence with LSTMs (Seq2Seq), Transformer, autoregressive fully connected neural nets (FC), Neural ODE, graph convolution networks (GCN) and graph attention networks (GAN). To train these DL models, we standardize I , R and D time series of each state individually to avoid one set of features dominating another. We use sliding windows to generate samples of sequences before the week that we want to predict and split them into training and validation sets. To train ODEs-based models, we rescale the trajectories of the number of cumulative cases of each state by the population of that state. We perform exhaustive search of the hyperparameters, including the learning rate, hidden dimensions and number of layers, for every DL model on the validation set. All these DL models are trained to predict the number of daily new cases instead of the number of accumulated cases because we want to detrend the time series, and put the training and test samples in the same approximate range. For graphical models, we view each state as a node, and then the adjacency matrix is the US states adjacency matrix. All experiments were run on Amazon Sagemaker (Liberty et al., 2020).

Results Table 1 shows the 7-day ahead prediction mean absolute errors of three features I , R and D for the weeks of July 13, Aug 23 and week Sept 6. We can see that AutoODE overall performs better than SuEIR and all the DL models. FC and Seq2Seq have better prediction accuracy of death counts. All DL models have much bigger errors on the prediction of week July 13, which may be due to insufficient training data. Another reason is that the number of cases in most states increase dramatically in July, and the test data is outside of the training data range, and neural networks are known to not be reliable in these cases (Kouw and Loog, 2018; Amodei et al., 2019). Figure 1 shows the 7-day ahead COVID-19 predictions of I , R and D in Massachusetts by AutoODE and the best DL model (FC). The prediction by AutoODE is closer to the target and has smaller confidence intervals. This demonstrates the effectiveness of our model, and the benefits of the combination of machine learning techniques with compartmental models.

MAE	07/13 ~ 07/19			08/23 ~ 08/29			09/06 ~ 09/12		
	I	R	D	I	R	D	I	R	D
FC	8379	5330	257	559	701	30	775	654	33
Seq2Seq	5172	2790	99	781	700	40	728	787	35
Transformer	8225	2937	2546	1282	1308	46	1301	1253	41
NeuralODE	7283	5371	173	682	661	43	858	791	35
GCN	6843	3107	266	1066	923	55	1605	984	44
GAN	4155	2067	153	1003	898	51	1065	833	40
SuEIR	1746	1984	136	639	778	39	888	637	47
AutoODE	818	1079	109	514	538	41	600	599	39

Table 1: Proposed AutoODE wins in predicting I and R : 7-day ahead prediction MAEs on COVID-19 trajectories of accumulated number of infectious, removed and death cases.

5 Conclusion

We study the problem of forecasting COVID-19 dynamics among the number of infected, removed and death cases in each US state. We perform a comprehensive benchmarking study of deep learning models, and present our physics-based approach, AutoODE, that has the best overall prediction performance, and obtains 57.4% reduction in mean absolute errors compared with the best competitor and 36.5% reduction compared with SuEIR. We find that FC and Seq2Seq have better prediction accuracy on the number of deaths, while the DL methods perform significantly worse on the prediction of the number of infected and removed cases than the physics-based models.

References

- Al-Aradi, A., Correia, A., Naiff, D., Jardim, G., and Saporito, Y. (2018). Solving nonlinear and high-dimensional partial differential equations via deep learning. *arXiv preprint arXiv:1811.08782*.
- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2019). Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*.
- Anderson, R. M., Anderson, B., and May, R. M. (1992). *Infectious diseases of humans: dynamics and control*. Oxford university press.
- Ayed, I., Bézenac, E. D., Pajot, A., and Gallinari, P. (2019). Learning partially observed PDE dynamics with neural networks.
- Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, B., Maddix, D. C., Türkmen, A., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Callot, L., and Januschowski, T. (2020). Neural forecasting: Introduction and literature overview. *ArXiv*, abs/2004.10240.
- Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. (2018). Neural ordinary differential equations. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 6571–6583. Curran Associates, Inc.
- Chen, Y.-C., Lu, P.-E., Chang, C.-S., and Liu, T.-H. (2020). A time-dependent sir model for covid-19 with undetectable infected persons. *arXiv preprint arXiv:2003.00122*.
- Davis, J. T., Chinazzi, M., Perra, N., Mu, K., y Piontti, A. P., Ajelli, M., Dean, N. E., Gioannini, C., Litvinova, M., Merler, S., Rossi, L., Sun, K., Xiong, X., Halloran, M. E., Jr., I. M. L., Viboud, C., and Vespignani, A. (2020). Estimating the establishment of local transmission and the cryptic phase of the covid-19 pandemic in the usa. *medRxiv preprint <https://doi.org/10.1101/2020.07.06.20140285>*.
- Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time. *Lancet Inf Dis.*, 20(5):533–534. doi:10.1016/S1473–3099(20)30120–1.
- Flunkert, V., Salinas, D., and Gasthaus, J. (2017). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *ArXiv*, abs/1704.04110.
- J.C.Butcher (1996). *Applied Numerical Mathematics*, volume 20. Elsevier B.V.
- Keskinocak, P., Aglar, B. E. O., Baxter, A., Asplund, J., and Serban, N. (2020). The impact of social distancing on covid19 spread: State of georgia case study. *medRxiv preprint <https://doi.org/10.1101/2020.04.29.20084764>*.
- Kouw, W. M. and Loog, M. (2018). An introduction to domain adaptation and transfer learning. *arXiv preprint arXiv:1812.11806*.
- Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., and Yan, X. (2020). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *arXiv preprint arXiv:1907.00235*.
- Liberty, E., Karnin, Z., Xiang, B., Rouesnel, L., Coskun, B., Nallapati, R., Delgado, J., Sadoughi, A., Astashonok, Y., Das, P., Balioglu, C., Chakravarty, S., Jha, M., Gautier, P., Arpin, D., Januschowski, T., Flunkert, V., Wang, Y., Gasthaus, J., Stella, L., Rangapuram, S., Salinas, D., Schelker, S., and Smola, A. (2020). Elastic machine learning algorithms in amazon sagemaker. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, SIGMOD '20*, page 731–737, New York, NY, USA. Association for Computing Machinery.
- Lim, B. and Zohren, S. (2020). Time series forecasting with deep learning: A survey. *ArXiv*, abs/2004.13408.
- Pei, S. and Shaman, J. (2020). Initial simulation of sars-cov2 spread and intervention effects in the continental us. *medRxiv preprint <https://doi.org/10.1101/2020.03.21.20040303>*.

- Raissi, M. and Karniadakis, G. E. (2018). Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141.
- Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., and Januschowski, T. (2018). Deep state space models for time series forecasting. In *NeurIPS*.
- Sirignano, J. and Spiliopoulos, K. (2018). Dgm: A deep learning algorithm for solving partial differential equations. *arXiv preprint arXiv:1708.07469*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *ArXiv*.
- Velickovic, P., mCucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Wang, Q., Xie, S., Wang, Y., and Zeng, D. (2020a). Survival-convolution models for predicting covid-19 cases and assessing effects of mitigation strategies. *medRXiv preprint <https://doi.org/10.1101/2020.04.16.20067306>*.
- Wang, R., Kashinath, K., Mustafa, M., Albert, A., and Yu, R. (2020b). Towards physics-informed deep learning for turbulent flow prediction. *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery and data mining*.
- Wen, R., Torkkola, K., Narayanaswamy, B., and Madeka, D. (2018). A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv:1711.11053*.
- Wu, N., Green, B., Ben, X., and O’Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317*.
- Zou, D., Wang, L., Xu, P., Chen, J., Zhang, W., and Gu, Q. (2020). Epidemic model guided machine learning for covid-19 forecasts in the united states. *medRXiv preprint <https://doi.org/10.1101/2020.05.24.20111989>*.

A Numerical Ordinary Differential Equation Solvers

Numerical methods are used to solve these coupled ordinary differential equations. Suppose we have an ordinary equation with initial condition,

$$\frac{dy}{dt} = f(t; y); \quad y(t_0) = y_0; \quad (\text{A.1})$$

The simplest method to solve this is Euler Method, which a first-order numerical method. It assumes the derivative is constant over a short period to approximate solution.

$$y_{n+1} = y_n + \Delta t f(t_n; y_n); \quad (\text{A.2})$$

where Δt is the time step.

A more accurate method is Runge-Kutta Method. The following equations shows the 4th order Runge-Kutta Method (RK4). k_1 is the slope at the beginning of the interval, if we only use k_1 , then it is just Euler Method. Both k_2 and k_3 are the slope at the midpoint of the interval, and k_4 is the slope at the end of interval.

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6} f(t_n; y_n); \\ k_1 &= f(t_n; y_n); \\ k_2 &= f\left(t_n + \frac{\Delta t}{2}; y_n + \Delta t \frac{k_1}{2}\right); \\ k_3 &= f\left(t_n + \frac{\Delta t}{2}; y_n + \Delta t \frac{k_2}{2}\right); \\ k_4 &= f\left(t_n + \Delta t; y_n + \Delta t k_3\right); \end{aligned} \quad (\text{A.3})$$

B Learn synthetic Correlated SuEIR Data

We generate synthetic 50 correlated time series of length 60 based on Equ 3.1 with the fourth order Runge-Kutta Method. All the parameters in Equ 3.1 are randomly generated. We tried training our model AdjMask-SuEIR given first 10 and 30 steps as input to predict the rest of time series only based on the loss l , R and D . From Figure 2, we can see that the model is able to correctly learn the trajectories of all six variables given first 30 steps as input. Since there are so many trainable parameters, so we show the mean absolute error between the true parameters and the learned parameters in Table 2. The MAEs of all parameters except for β and A are small because the solutions are not unique.

MAE	Learned (30)	Learned (10)
β	0.4635	0.3761
A	0.3256	0.3286
γ	0.0014	0.0003
μ	0.0761	0.108
σ	0.0878	0.1213
E_0	0.1511	0.1782
S_0	0.1511	0.1873

Table 2: The mean absolute error between the true parameters and the learned parameters given 30, 10 steps of training data

C Additional Figures

