SAND2020-11933R

Physics-Informed Machine Learning for Epidemiological Models

Carianne Martinez, Jessica Jones, Drew Levin, Nat Trask, Patrick Finley

September 2020

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Funding Acknowledgment
Research was supported by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act.

Disclaimer
Unless otherwise indicated, this information has been authored by an employee or employees of National Technology and Engineering Solutions of Sandia, LLC, operator of Sandia National Laboratories with the U.S. Department of Energy. The U.S. Government has rights to use, reproduce, and distribute this information. The public may copy and use this information without charge, provided that this Notice and any statement of authorship are reproduced on all copies.
While every effort has been made to produce valid data, by using this data, User acknowledges that neither the Government nor operating contractors of the above national laboratories makes any warranty, express or implied, of either the accuracy or completeness of this information or assumes any liability or responsibility for the use of this information. Additionally, this information is provided solely for research purposes and is not provided for purposes of offering medical advice. Accordingly, the U.S. Government and operating contractors of the above national laboratories are not to be liable to any user for any loss or damage, whether in contract, tort (including negligence), breach of statutory duty, or otherwise, even if foreseeable, arising under or in connection with use of or reliance on the content displayed in this report.

1. GOAL

One challenge of using compartmental SEIR models for public health planning is the difficulty in manually tuning parameters to capture behavior reflected in the real-world data.

This team conducted initial, exploratory analysis of a novel technique to use physics-informed machine learning tools to rapidly develop data-driven models for physical systems. This machine learning approach may be used to perform data assimilation of compartment models which account for unknown interactions between geospatial domains (i.e. diffusion processes coupling across neighborhoods/counties/states/etc.).

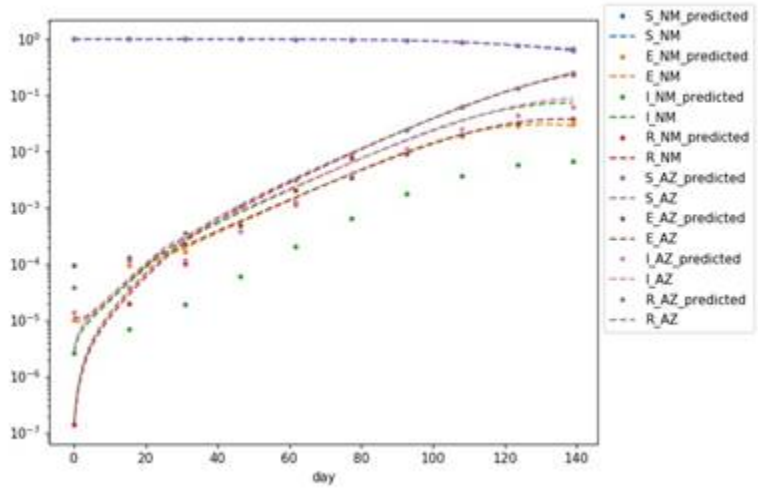
Results presented here are early, proof-of-concept ideas that demonstrate initial success in using a physically informed neural network (PINN) model to assimilate data in a compartmental epidemiology model. The results demonstrate initial success and warrant further research and development.

2. INITIAL RESULTS

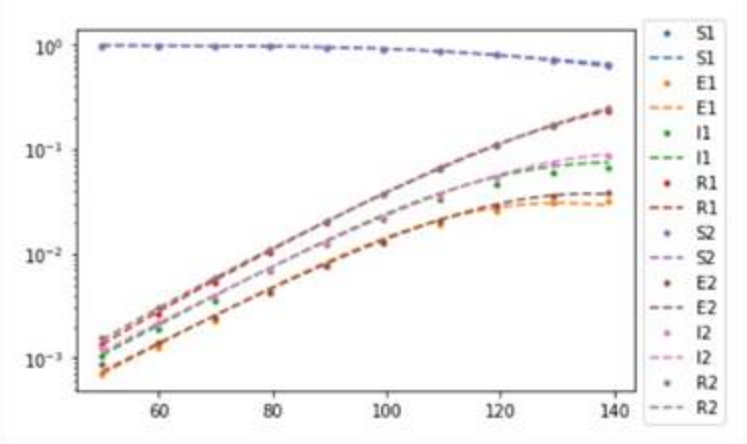
ODE models representing COVID-19 epidemiological infection dynamics for New Mexico and Arizona were prepared and parameters were fit using actual data.

We pick 10 points in time, equally spaced, and determine the model's loss by comparing the prediction vs the actual at those 10 points. We use those 10 points to compute the derivatives for the ODE calculations (recognizing that 10 points is not a lot to support a numerical integration of an ODE system of any complexity). At first, we assume the parameter values are constant throughout the simulation, which is unlikely; the beta parameter that controls viral infectivity could change drastically over time as people adopt or abandon mitigation strategies such as social distancing and mask wearing.

We then fit the model over several 100-day windows to see which windows produce the best fit. This results in a few simulations with decent fits, but due to the dependence of the parameters to the chosen window, they offer poor predictive value.

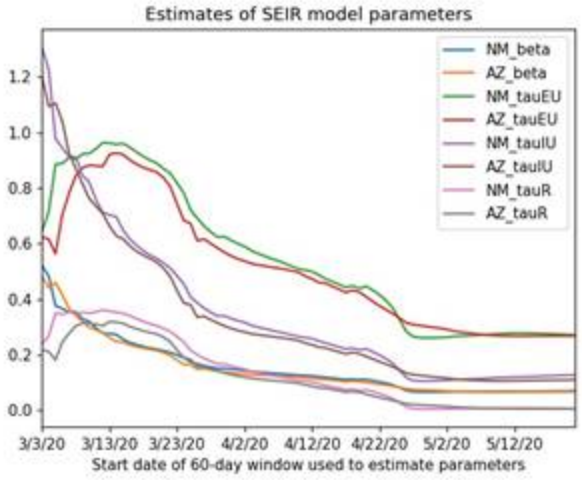


Some of the windows produce fits that are potentially acceptable. Here is an example of a window that results in a good fit, but the fits cannot necessarily be extrapolated to future time points, nor do the parameter values found here represent dynamics that occur outside the window.



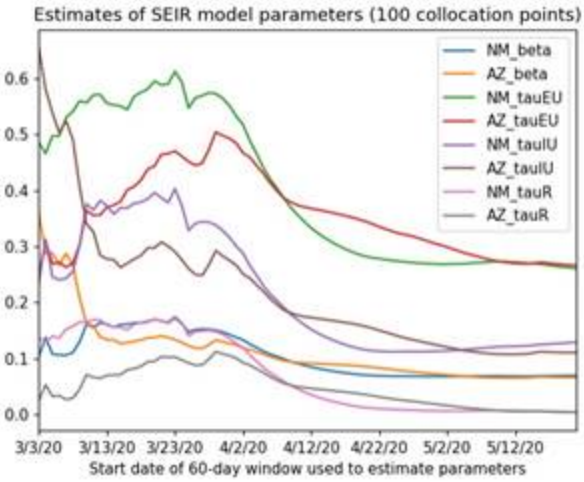

```
Beta estimates [0.07546327137382697, 0.06705320804096054]
```

We then refit the model based on sliding 60-day windows to see how the NM and AZ values of Beta (viral infectivity) and Tau (inverse of the expected infection time) vary over time.



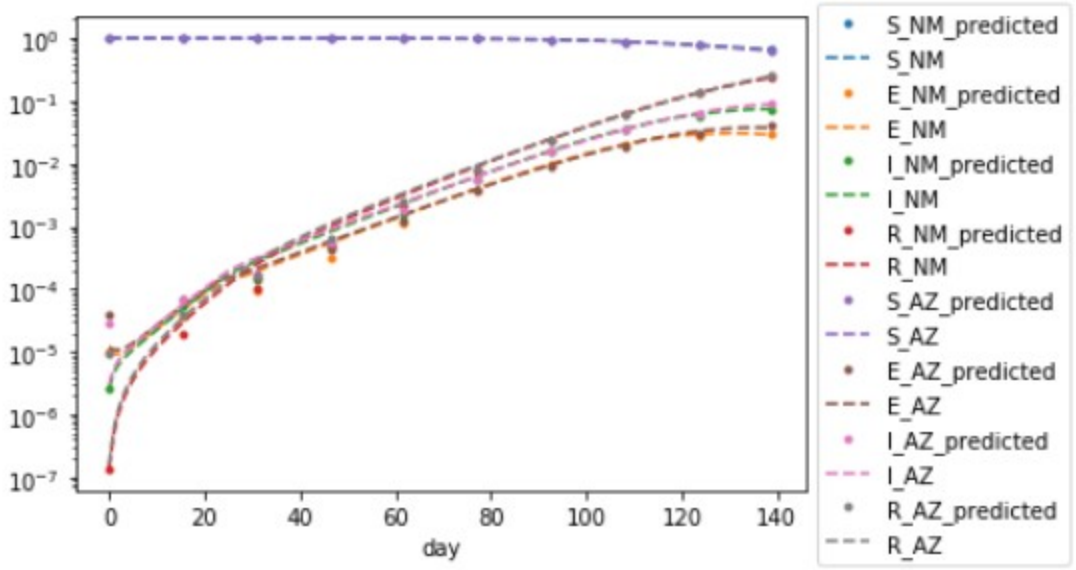
The results suggest high infectivity earlier on, though if anything this suggests that a model with static parameters is not sufficient to capture real-world infection dynamics.

We then modified the model to increase the number of calculated points from 10 to 100. This has the dual benefit of allowing the model to be more sensitive in its loss function and be more accurate when calculating derivatives for numerical integration.



The parameter values now fall more clearly into two 'phases', possibly before and after social distancing, though these models are all fit to 60-day windows still assuming static parameter values.

Finally, we extended the model to support time-dependent values for the parameters. Allowing Beta (viral infectivity) to vary over time significantly improved fits when applied to the full dataset:



The Beta parameter values over time in this simulation are:

NM: [0.0527, 0.0653, 0.0735, 0.0768, 0.0754, 0.0706, 0.0642, 0.0584, 0.0550, 0.0550]

AZ: [0.0455, 0.0598, 0.0689, 0.0727, 0.0718, 0.0679, 0.0636, 0.0612, 0.0626, 0.0689]

The results show an increase in infectivity early followed by a return to lower levels by the end of the simulation.

Our effort shows that extending the ODE model and using a PINN to fit the data to parameters that are allowed to vary over time, result in significantly improved model fits. Significantly more work is required to extend this machine learning approach and offer improved epidemiological modeling advancements.