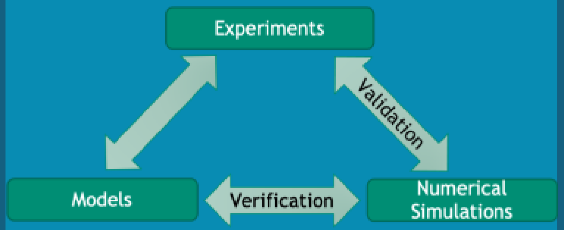
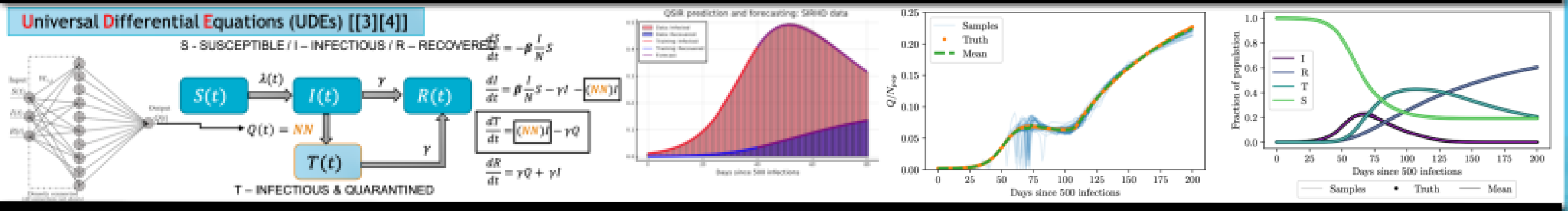
SAND2021-8359C

Sandia National Laboratories

Assessing the Efficacy of Universal Differential Equations to Learn Missing Dynamics from a Subset of Observable State Variables





Presented By

Erin C.S. Acquesta (5954) & Teresa Portone (1463)

In collaboration with:

Chris Rackauckas (MIT) & Raj Dandekar (MIT)

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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.

All Models are Wrong, Some are Useful

-George Box

Modeling real-world phenomena to any degree of accuracy is a challenge that the scientific research community has navigated since its foundation.

Historically researchers have utilized and continue to utilize:

Verification: Are we solving the equations correctly?
Validation: Are we solving the correct equations?
Uncertainty Quantification:

- What uncertainty is attributed to inherent random behavior (i.e. aleatory)?
- What uncertainty is attributed to the fact that we don't know (i.e. epistemic)?

System Identification: Statistical methods to build mathematical models of dynamical systems from measured data.

Today:

Scientific Machine Learning: Area of machine learning focused on the use of machine learned models used in lieu of, complementary to, or as surrogates for computational simulation models used for science and engineering.

Epidemiology

The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems [1]

Nomenclature:

Basic Reproduction Number, R_0, "R-naught"

The expected number of infections from one infected individual introduced into a population of 100% susceptible individuals.

Replacement Number, $R_{\text{eff}}(\text{t})$, "R-effective"

After the early stages of an epidemic has passed, the number of secondary infections is expected to go down as the number of susceptible individuals goes down.

Virus

Pathology

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

Viruses are named based on their genetic structure to facilitate the development of diagnostic tests, vaccines and medicines.

Disease

Epidemiology

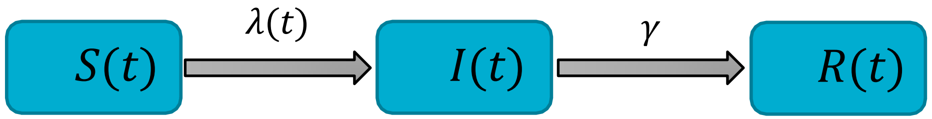
COronaVIrus Disease, 2019 (COVID-19)

Diseases are named to enable discussion on disease prevent, spread, transmissibility, severity and treatment.

For more details, please reference the CDC's *Introduction to Epidemiology*: https://www.cdc.gov/csels/dsepd/ss1978/Lesson1/Section1.html#_ref1

Compartmental Models for Infectious Disease [2]

S - SUSCEPTIBLE / I – INFECTIOUS / R – RECOVERED



System of ordinary differential equations (ODEs):

$$\frac{dS(t)}{dt} = -\beta \frac{I(t)}{N} S(t)$$

$$\frac{dI(t)}{dt} = \beta \frac{I(t)}{N} S(t) - \gamma I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t)$$

Force of Infection Function: $\lambda(t) := \beta \frac{I}{N}$

Infection rate: β

Recover rate: γ

Residence Time in $I(t)$: $\frac{1}{\gamma}$

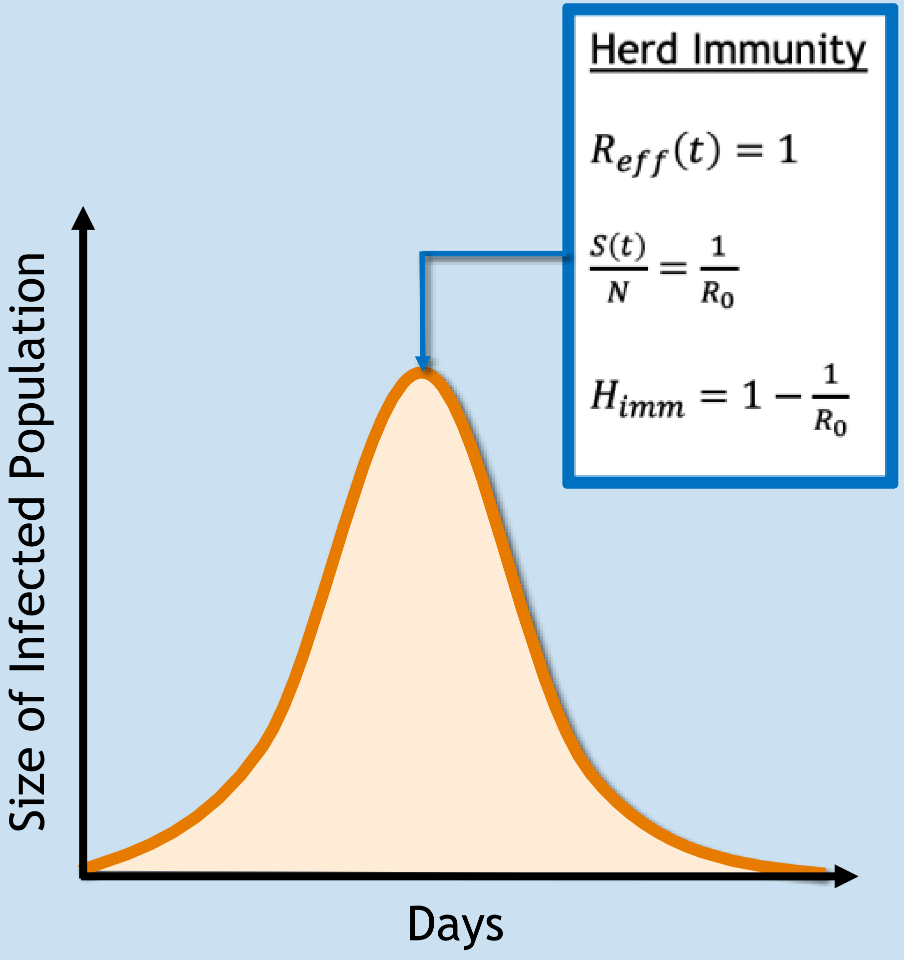
Total Population: $N = S(t) + I(t) + R(t)$

*Note: the total population size is constant under this formulation

Model $R_0 = \beta \frac{1}{\gamma}$

Model $R_{eff}(t) = R_0 \left(\frac{S(t)}{N}\right)$

Notional Plot of Infected Population, $I(t)$



Case Counts for Covid-19 in United States

We know the classic SIR model is under-representative of the real -world phenomenon it is intended to simulate.

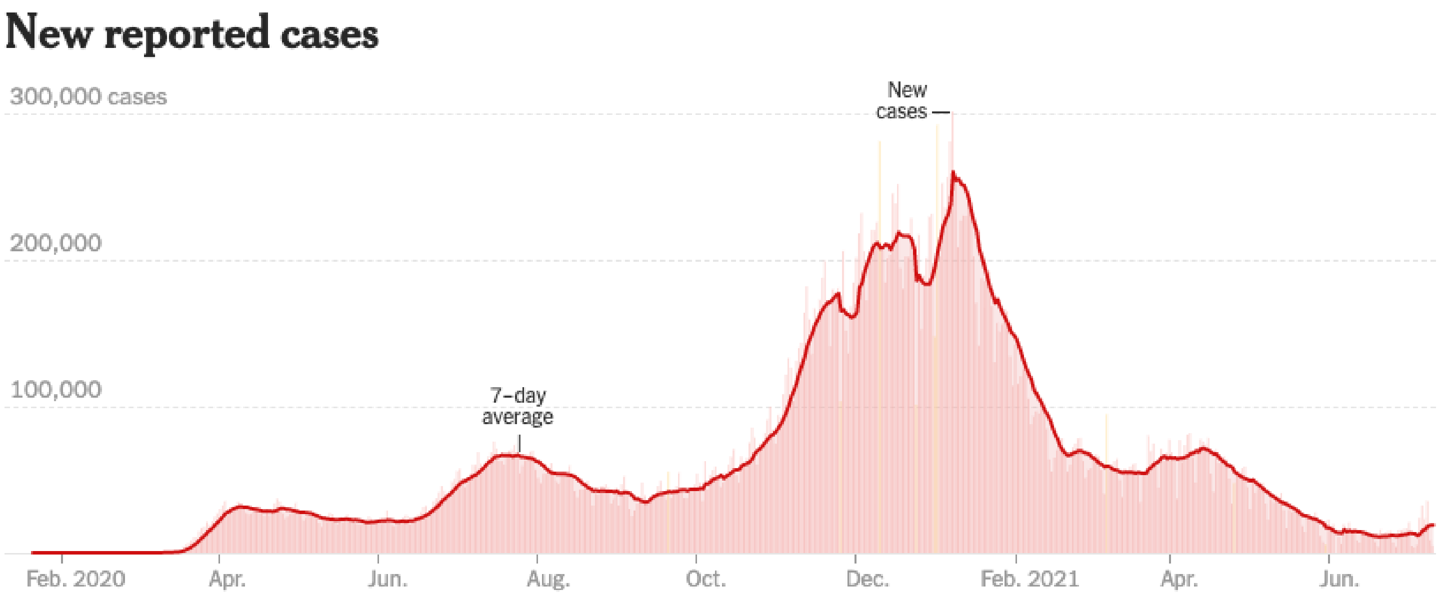


Image Credit: NYT https://www.nytimes.com/interactive/2021/us/covid-cases.html [accessed 2021/07/12]

"All models are wrong":

- Homogenous behaviors.
- Uniform mixing of the population.
- Lacks influence on disease transmission from interventions.
  - Quarantine
  - Social Distancing
  - Personal Protective Equipment (PPE)

"Some are useful":

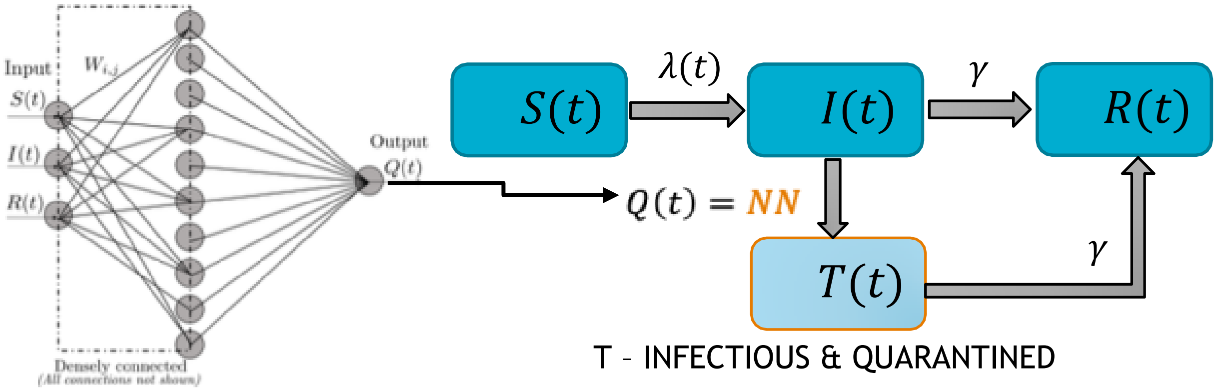
- During an emergent pandemic, we do NOT have the details to build heterogeneous models.
- Model calibration may provide insight into the baseline parameters.
- These are the baseline dynamics that have been used for decades in all infectious disease models.

Scientific Machine Learning (SciML)

Area of machine learning focused on the use of machine learned models used in lieu of, complementary to, or as surrogates for computational simulation models used for science and engineering.

Universal Differential Equations (UDEs) [[3][4]]

S - SUSCEPTIBLE / I – INFECTIOUS / R – RECOVERED



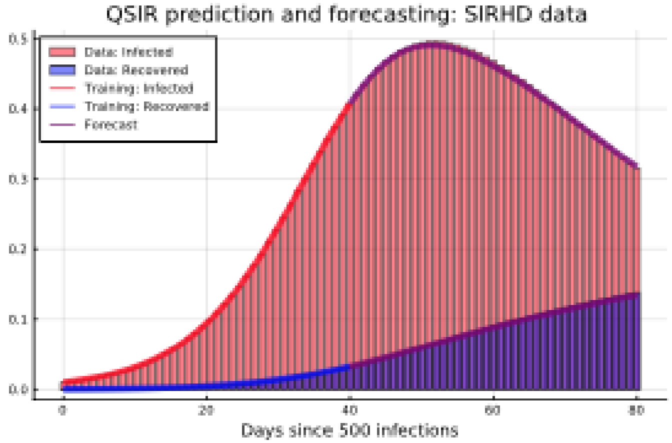
T – INFECTIOUS & QUARANTINED

$$\frac{dS}{dt} = -\boldsymbol{\beta}\frac{I}{N}S$$

$$\frac{dI}{dt} = \boldsymbol{\beta}\frac{I}{N}S - \gamma I - (NN)I$$

$$\frac{dT}{dt} = (NN)I - \gamma Q$$

$$\frac{dR}{dt} = \gamma Q + \gamma I$$



Physics-Informed Neural Networks (PINNS) [5]

Data-driven solutions to Partial Differential Equations (PDEs)

$$u_t + \mathcal{N}[u] = 0, \qquad x \in \Omega \subset \mathbb{R}^m, t \in [0, T]$$

where $u(t,x)$ denotes the latent (hidden) solution, $\mathcal{N}[\cdot]$ is a nonlinear differential operator

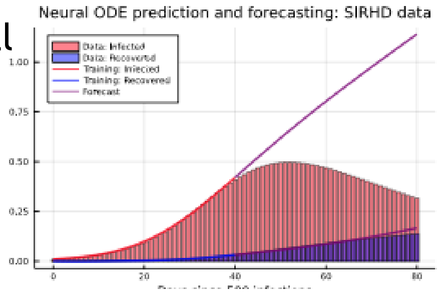
Then…. $u(t,x) = NN$

Neural Ordinary Differential Equations (ODEs) [6]

Simulating unknown dynamics for a full System of ODEs:

$$\frac{dU(t)}{dt} = NN(U)$$

where $U = [S, I, R]$



Universal Differential Equations (UDEs) [7]

UDEs are part of the SciML libraries written in Julia: https://sciml.ai
Julia is a high-level high-performance parallel computing language: https://julialang.org

Efficient Training of Universal Differential Equations via Differentiable Programming

- Given discrete data points: $(t_i, \boldsymbol{d}_i)$ where
  - $t_i \in [t_0, t_1]$ are discrete points along the time horizon and
  - $\boldsymbol{d}_i$ are the corresponding realizations representing the state solutions (e.g. $S(t), I(t), R(t)$).

- **Objective:** minimize the cost function, $C(\theta)$, on the current approximation to the solution, $u^{[j]}(t)$, to the dynamical system, $\frac{du}{dt} = F\big(u, t, NN_\theta(u,t)\big)$.

$$C(\theta) = \sum_i \left\| u^{[j]}(t_i) - \boldsymbol{d}_i \right\|$$

- "Differentiable programming framework with reverse-mode accumulation is used to allow for deriving on-the-fly approximation for the wide range of differential equation types."

Using pullbacks:

Given a function $f(x) = y$ the pullback at x is the function:

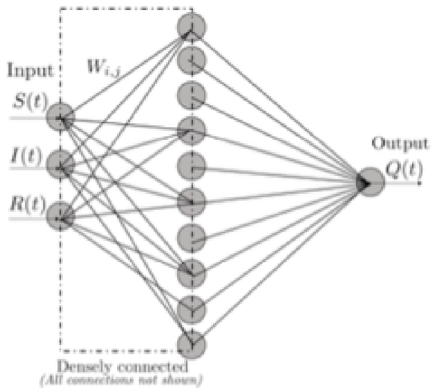
$$B_f^x(y) = y^T f'(x)$$

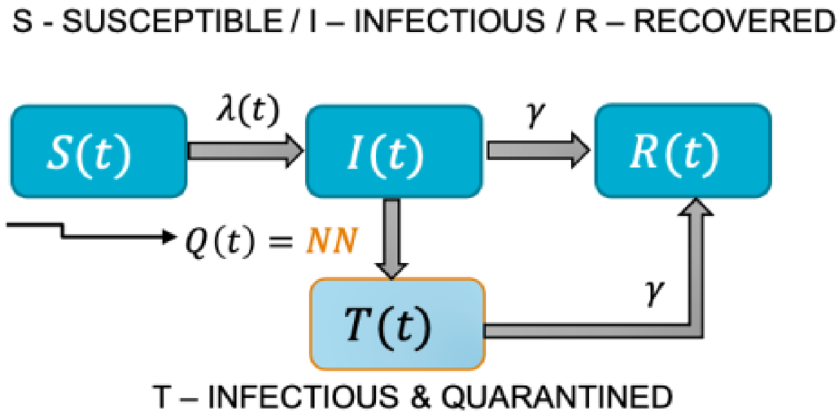
then, the pullback of a cost function computes the gradient.

Assessing the Effects of Quarantine Control in COVID-19 Spread [4]

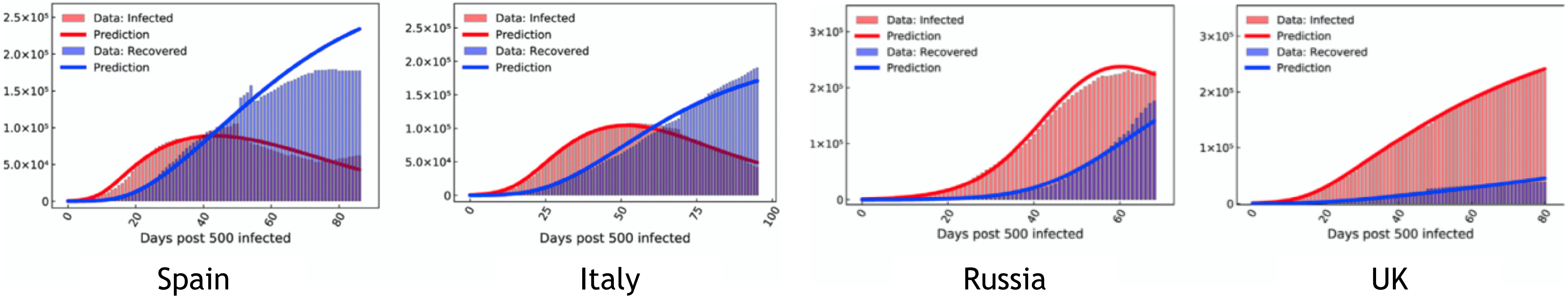
UDEs are a useful and valuable tool to learn what we don't already know!!

It can provide invaluable insight during an emergent pandemic.





Learning the effect of quarantine around the world:



UDEs for Compartment Models of Infectious Disease [4]

Dandekar et. al. used UDEs to represent isolation dynamics for COVID-19.

$$\frac{dS(t)}{dt} = -\frac{\beta I(t)}{N_{pop}} S(t)$$

$$\frac{dI(t)}{dt} = \frac{\beta I(t)}{N_{pop}} S(t) - \gamma I(t) - Q(t) I(t)$$

$$\frac{dT(t)}{dt} = Q(t) I(t) - \delta T(t)$$

$$\frac{dR(t)}{dt} = \delta I(t) + \delta T(t)$$

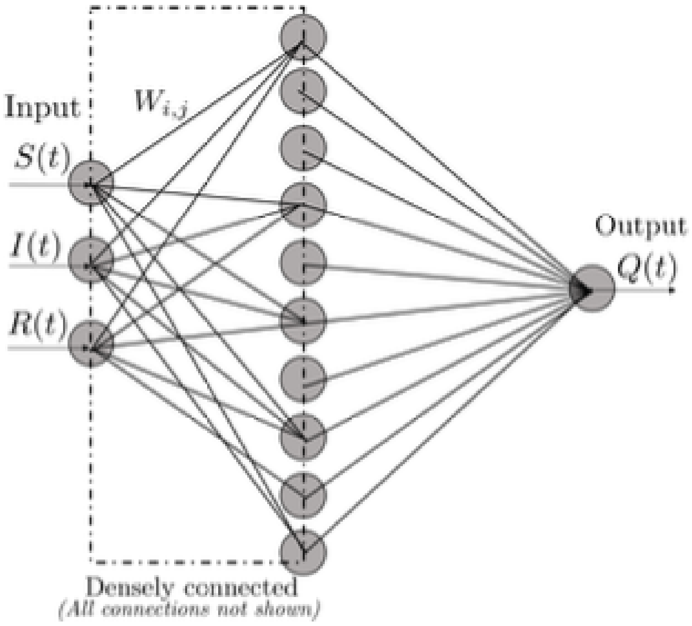
- Introduced an isolation state $T(t)$.
- UDE for the *nonlinear* transition rate from $I(t)$ into $T(t)$, denoted $Q(t)$.
- $Q(t)$ is approximated with a small neural network trained on data for $I(t)$ & $R(t)$, denoted $NN_\theta(u, t)$.
- By definition constrained to conserve population,

$$\frac{dN_{pop}}{dt} = 0$$

Loss Function:

$$L_{NN}(\theta, \beta, \gamma, \delta) = \left\|\log(I(t)) - \log(I_{data}(t))\right\|^2 + \left\|\log(R(t)) - \log(R_{data}(t))\right\|^2$$

Universal Approximation Theorem



Universal Approximation Theorem [[8],[9],[10]]:
(one version) Fix a continuous function $\sigma: \mathbb{R} \to \mathbb{R}$ (activation function) and positive integers d, D. The function σ is not a polynomial if and only if, for every continuous function $f: \mathbb{R}^d \to \mathbb{R}^D$ (target function), every compact set K of $\mathbb{R}^d$, and every $\varepsilon > 0$ there exists a continuous function $f_\varepsilon: \mathbb{R}^d \to \mathbb{R}^D$ (the layer output) with representation

$$f_\varepsilon = W_2 \circ \sigma \circ W_1$$

where W_2, W_1 are composable affine maps and $\circ$ denotes component-wise composition, such that the approximation is bounded

$$\sup_{x \in K} ||f(x) - f_\varepsilon(x)|| < \varepsilon$$

Although the Universal Approximation Theorem is necessary condition for neural networks to be universal approximators, it is not a sufficient condition in practice.

How do we know we are learning something useful?!?

Inferring Transition into Quarantine with Incomplete Data

- Dandekar et. al. used observations of $I(t), R(t)$ to infer transition rates (including $Q(t)$) for COVID-19.
- There are many applications for which *only a subset* of the state variables can be observed.

How does this affect the ability to recover "useful" information about $Q(t)$ & disease dynamics?!?

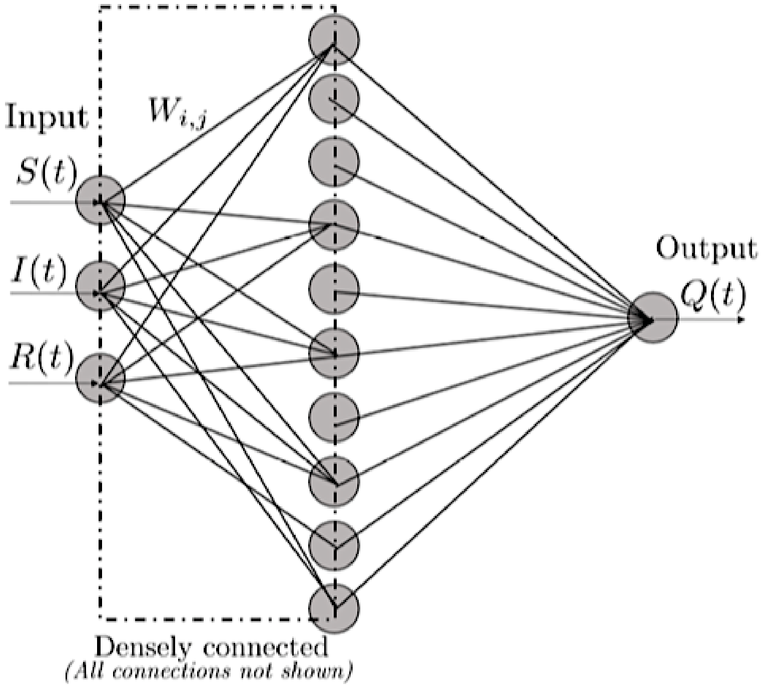
Experimental Plan

1. Generate synthetic data with prespecified NN_{θ^*}.
2. Infer $NN_{\hat{\theta}}$ & transition rate parameters $\hat{\beta}, \hat{\gamma}, \hat{\delta}$ from combinatorial subsets of state variable observations.
   a. Data = $[I, R, T], [I, R], [I, T], [R, T], [I], [R], [T]$
3. Study mean-squared error (MSE) of inferred $\hat{Q}(t)$ vs "true" $Q^*(t)$ used to generate data for each dataset to determine when inference degrades.

Problem Specification/State of Knowledge

Assumptions:

- NN_θ fully-connected neural network of depth 1, width 10, and a ReLU activation functions.
- NN_θ architecture the same for data generation & inference.
- Synthetic data not corrupted by noise.
- Initial condition assumed known for state variables $S(0), I(0), T(0),$ and $R(0)$.
- All transition rate parameters $\beta, \gamma,$ and δ uncertain; distributions derived from literature.
- Neural network parameters, $\boldsymbol{\theta}$ and ODE parameters, $\beta, \gamma,$ and δ



Ensemble Training: Robust Learning and Uncertainty Quantification

Approach

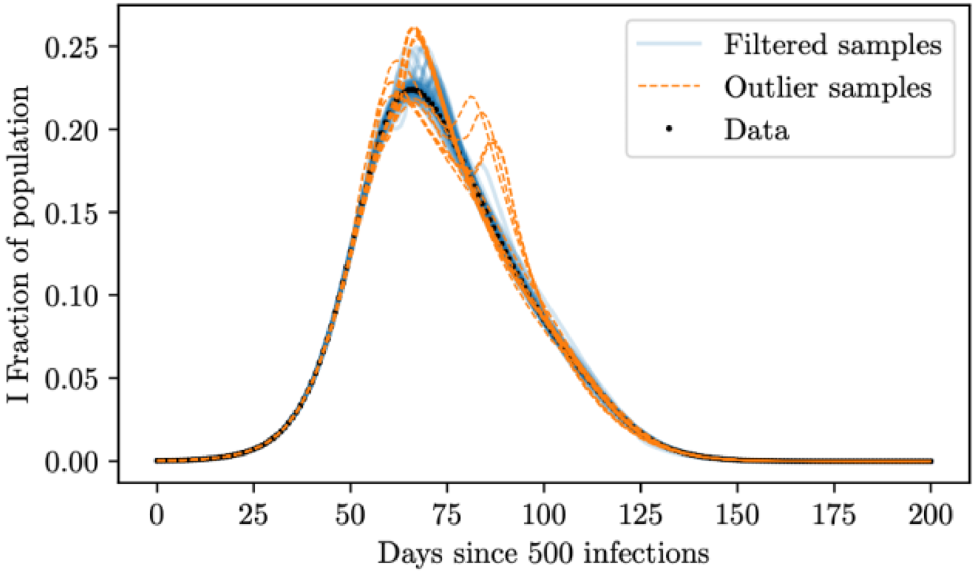
For each of the 7 possible observable state combinations: $\{[I,R,T],[I,R],[I,T],[R,T],[I],[R],[T]\}$

1. Run 100 training replicates to learn $\{NN_{\hat{\boldsymbol{\theta}}}^{k}\}$, $\{\hat{\beta}^k\}$, $\{\hat{\gamma}^k\}$, and $\{\hat{\delta}^k\}$, for $k = 1, \ldots, 100$.
   a. Each parameter is randomly initialized at the beginning of training.
   b. Due to the vanishing gradient problem:
      i. Filter out outlier ensemble members (those with very large MSE).
2. This will result in $n \leq 100$ realizations for optimal transition rate parameters and approximations for $Q(t)$

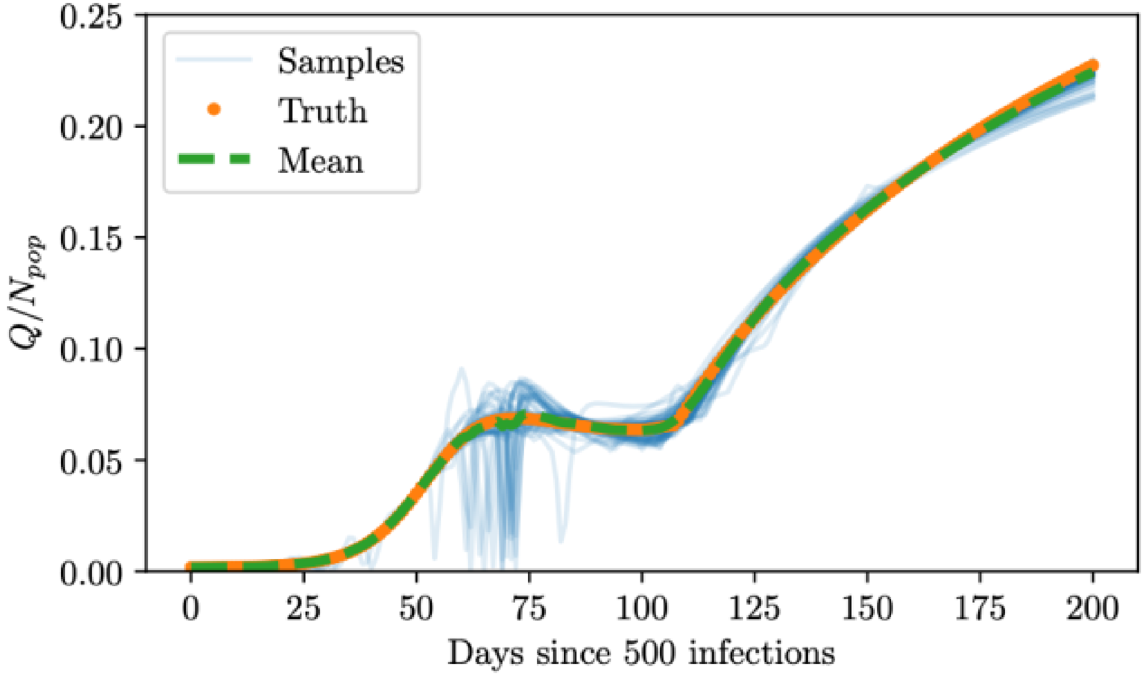
Uncertainty Quantification Opportunity

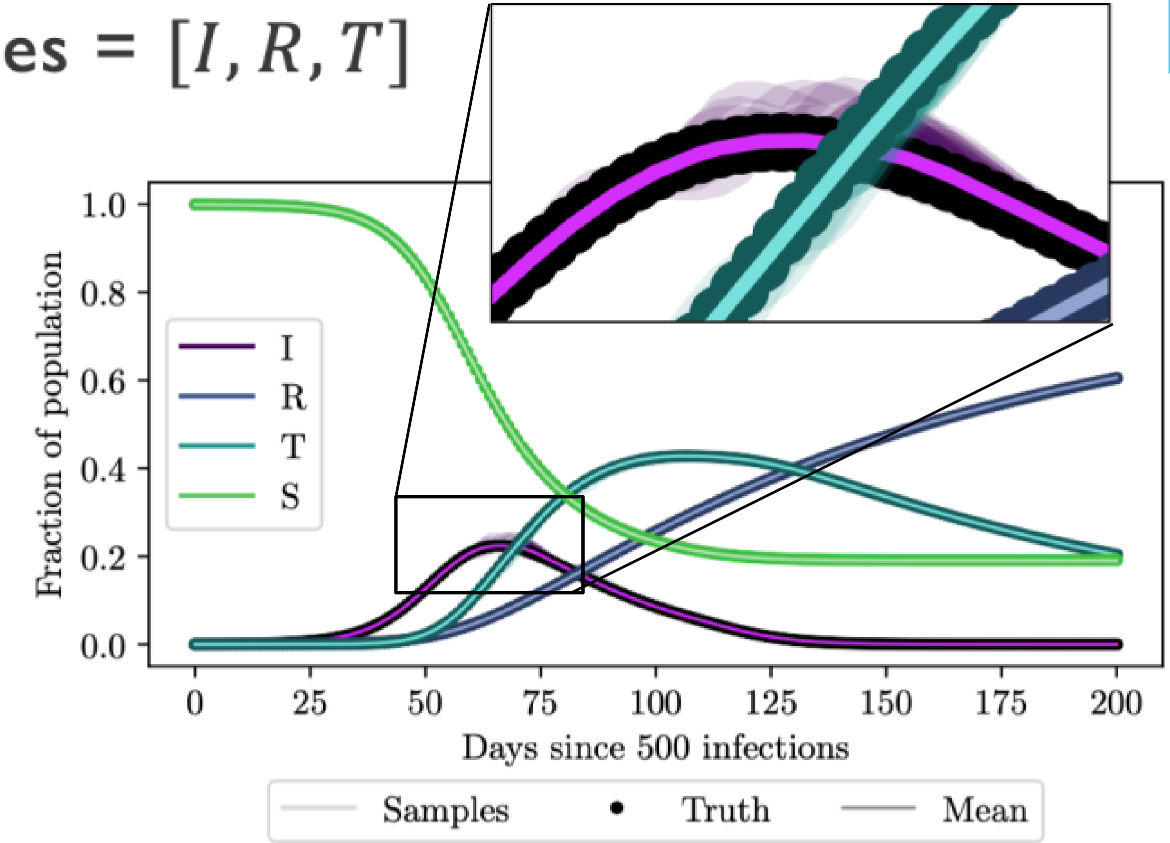
The ensembles provides an estimate on the variance for:

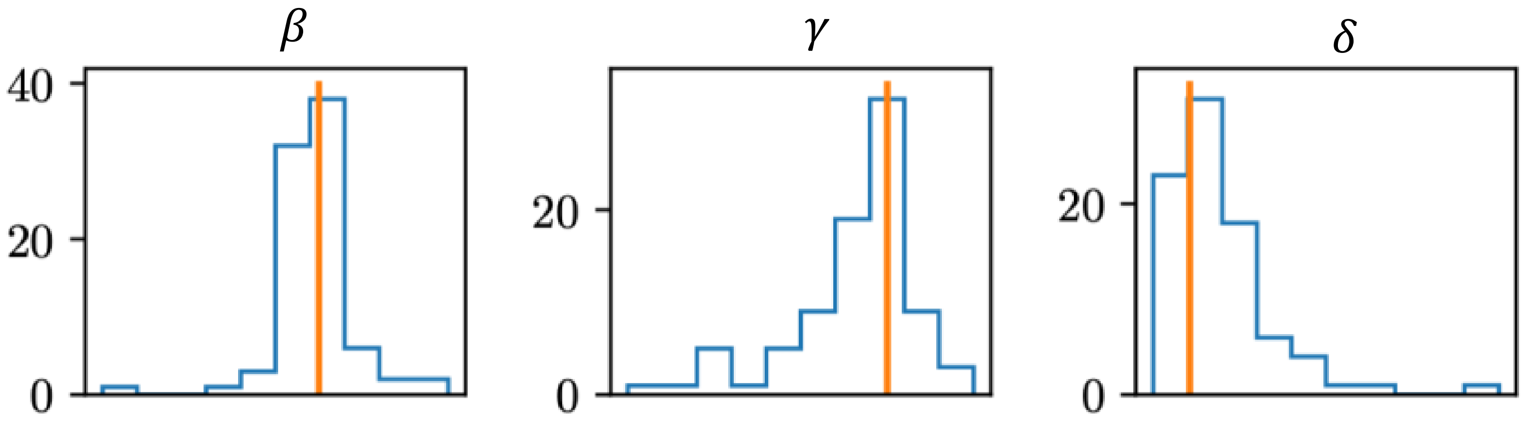
- Unobserved state variable trajectories.
- Optimized transition rate parameters.
- $Q(t)$



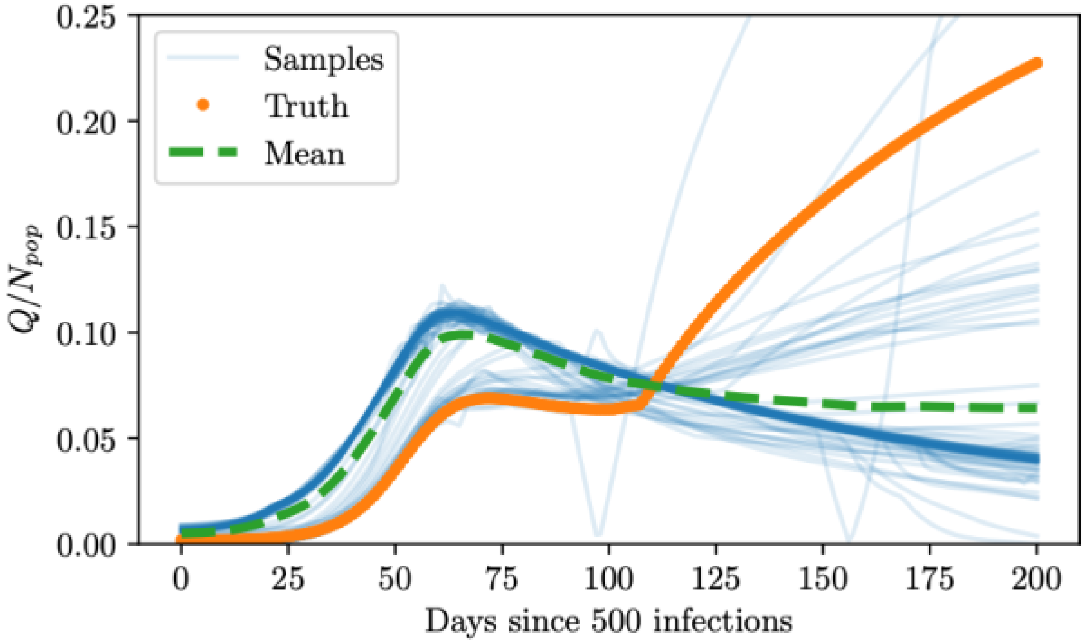
Training Results: Observable States = $[I, R, T]$

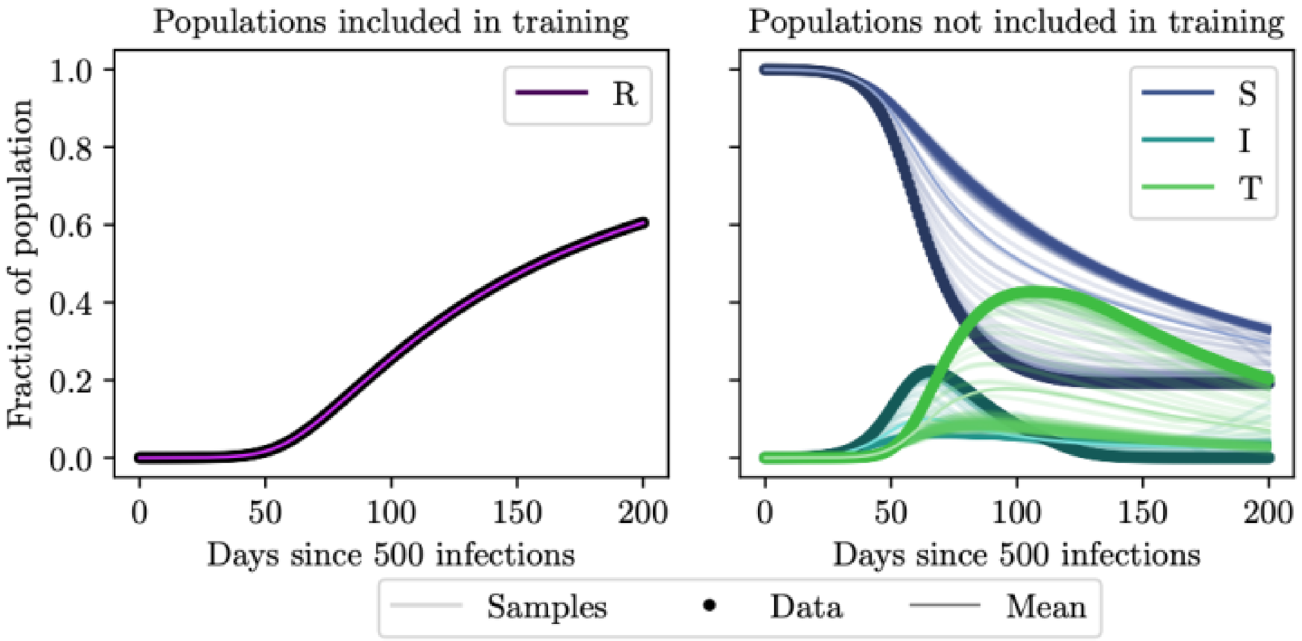


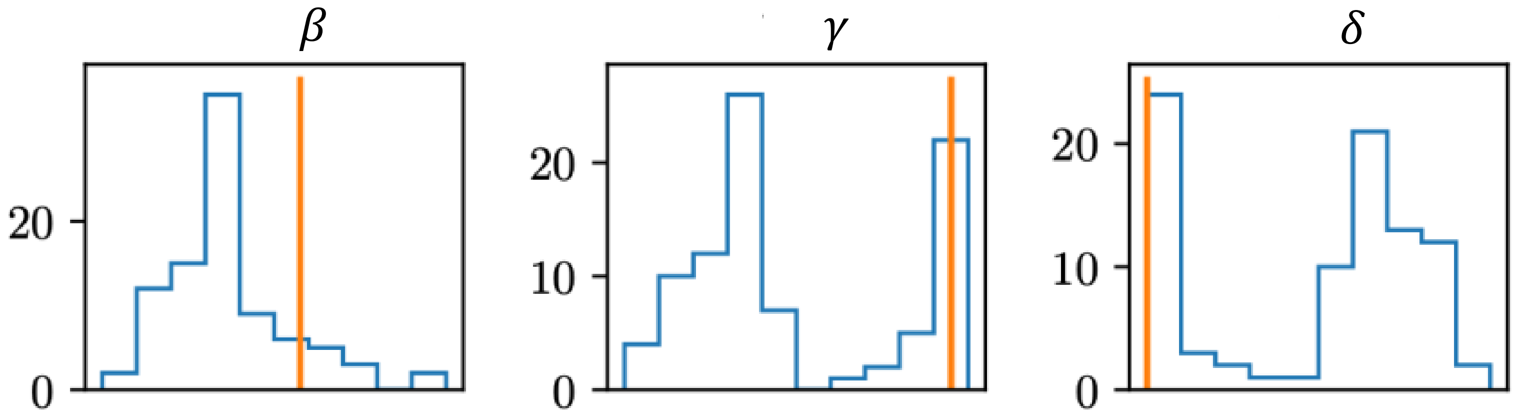




Training Results, Observable States = $[R]$





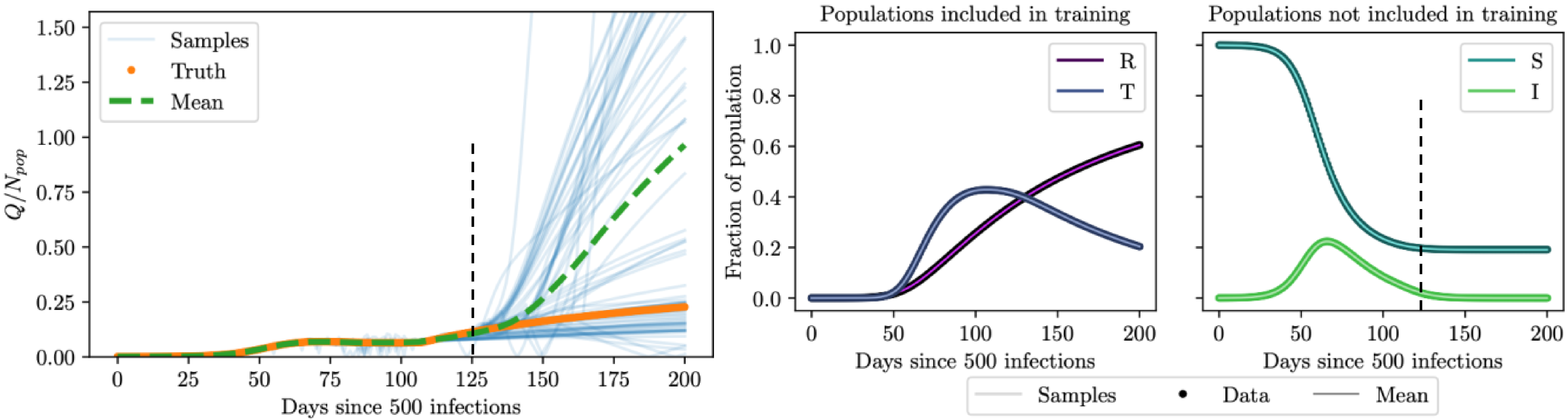


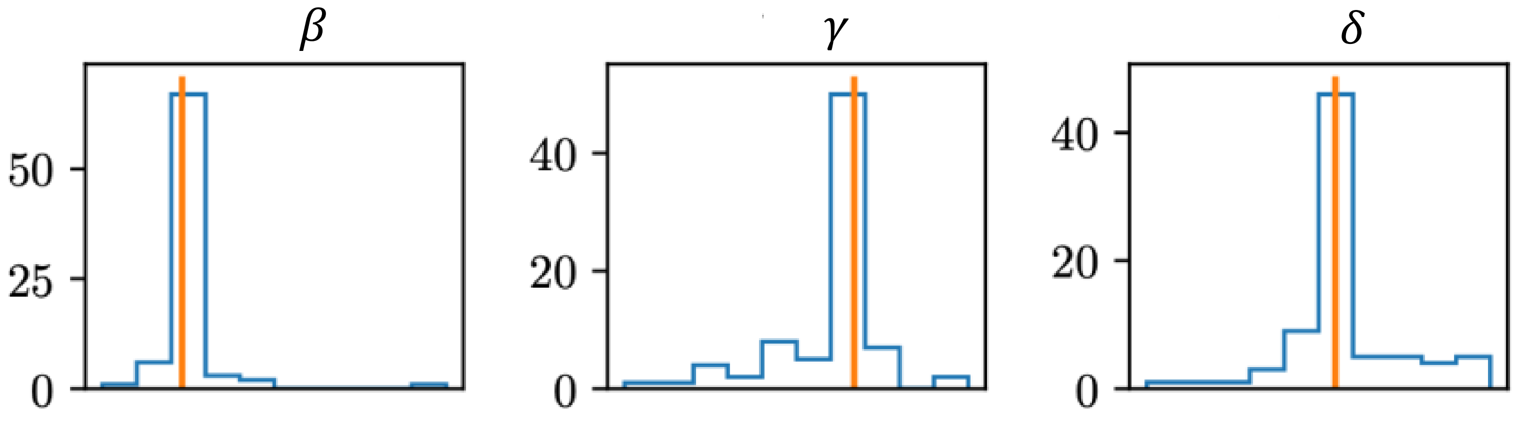
Ranking Q Recovery by Subset of Observable States

- Computed MSE of ensemble-mean (average) $\bar{Q}(t)$ vs "true" $Q^{\star}(t)$ used to generate the data.
- Ranked data scenarios by MSE.

Why is this the worst case?!?

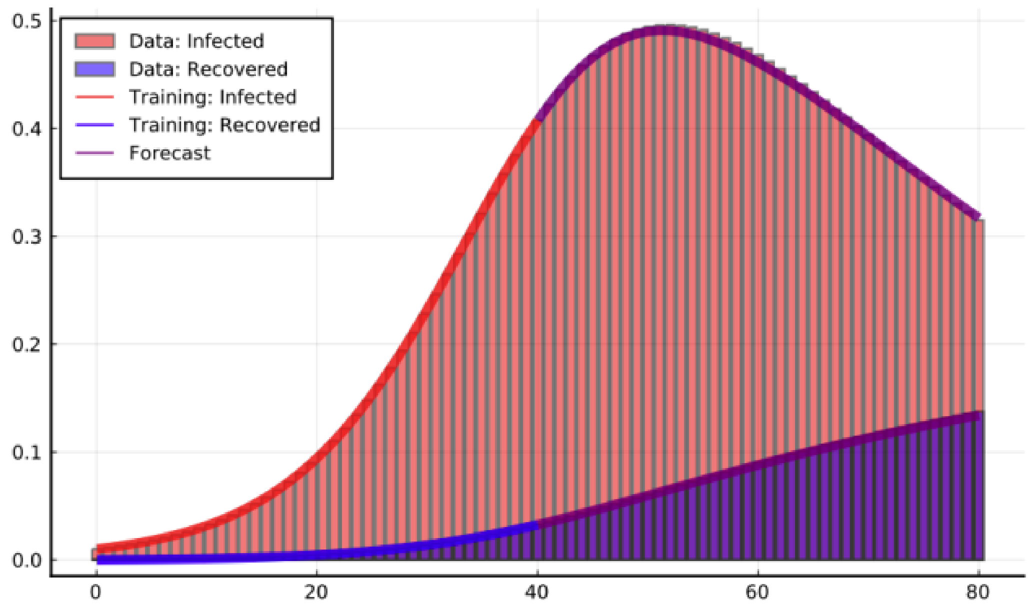
Training Results, Observable States = $[R, T]$





Conclusions & Future Work

- Developed a procedure to study success of UDE training when only able to observe subsets of state variables.
- Ensemble of training results provides understanding of uncertainty in inferred dynamics.
- Next steps:
  - Noisy and/or sparse data
  - Data generated from more complex model



- For more complex model must determine appropriate accuracy metric (no "true Q" to compare to).
- Potential metric: MSE of observed state variables extrapolated beyond time horizon of training data.

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.

References

1. Last, J.M. ed., 2001. *A dictionary of epidemiology* (p. 47). Oxford, UK: Oxford university press.
2. Hethcote, H.W., 2000. The mathematics of infectious diseases. *SIAM review*, *42*(4), pp.599-653.
3. Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D. and Ramadhan, A., 2020. Universal Differential Equations for Scientific Machine Learning. *arXiv preprint arXiv:2001.04385*.
4. Dandekar, R., Rackauckas, C. and Barbastathis, G., 2020. A machine learning-aided global diagnostic and comparative tool to assess effect of quarantine control in COVID-19 spread. *Patterns*, *1*(9), p.100145.
5. Raissi, M., Perdikaris, P. and Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, pp.686 -707.
6. Chen, R.T., Rubanova, Y., Bettencourt, J. and Duvenaud, D., 2018. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366*.
7. Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A. and Edelman, A., 2020. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*.
8. Leshno, M., Lin, V.Y., Pinkus, A. and Schocken, S., 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, *6*(6), pp.861-867.
9. Pinkus, A., 1999. Approximation theory of the MLP model. *Acta Numerica 1999: Volume 8*, *8*, pp.143-195.
10. Kratsios, A., 2019. Characterizing the Universal Approximation Property. *arXiv e-prints*, pp.arXiv-1910.
11. Smith, R.C., 2013. *Uncertainty quantification: theory, implementation, and applications* (Vol. 12). Siam.

Appendix

Initial sampling of transition rate params

- Transition rate parameters are interdependent.
- Instead randomly sample independent variables and computed derived transition rates.
- For SIRHDT model, e.g., hospitalized population H can flow into R or D.
- τ_{HR}, τ_{HD} can be defined in terms of
  - r_{HR}, r_{HD} ratios of populations flowing into H and D
  - T_H - residence time in hospitalized population
- Then

$$\tau_{HR} = \frac{r_{HR}}{T_H}, \qquad \tau_{HD} = \frac{r_{HD}}{T_H}$$

- Reasonable ranges to sample r_*, T_* derived from the literature and used to define distributions from which they are sampled for initial guesses to optimization.

Filtering procedure

- Computed MSE for each ensemble member, each population in the data.
- Filtered out any ensemble member whose MSE was deemed an outlier using the interquartile range (IQR) heuristic threshold:

$$Q_3 + 1.5\, IQR = Q_3 + 1.5(Q_3 - Q_1)$$

where Q_1= 0.25 quantile, Q_3=0.75 quantile.