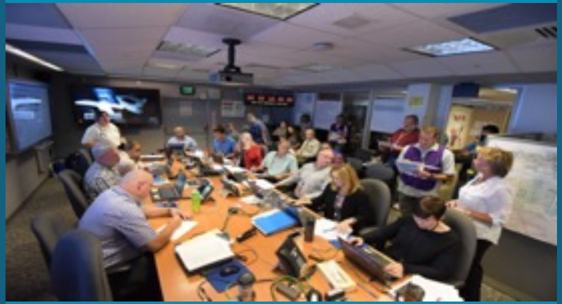
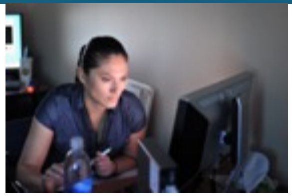
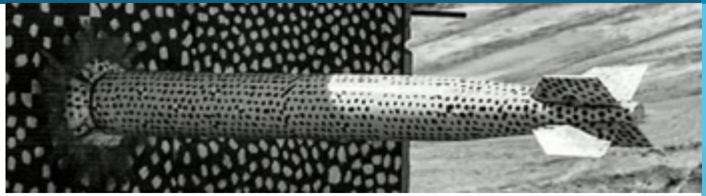
SAND2021-8583 C

Learning Missing Mechanisms in a Dynamical System from a Subset of State Variable Observations







Portone, T., Acquesta, E., Dandekar, R., Rackauckas, C.

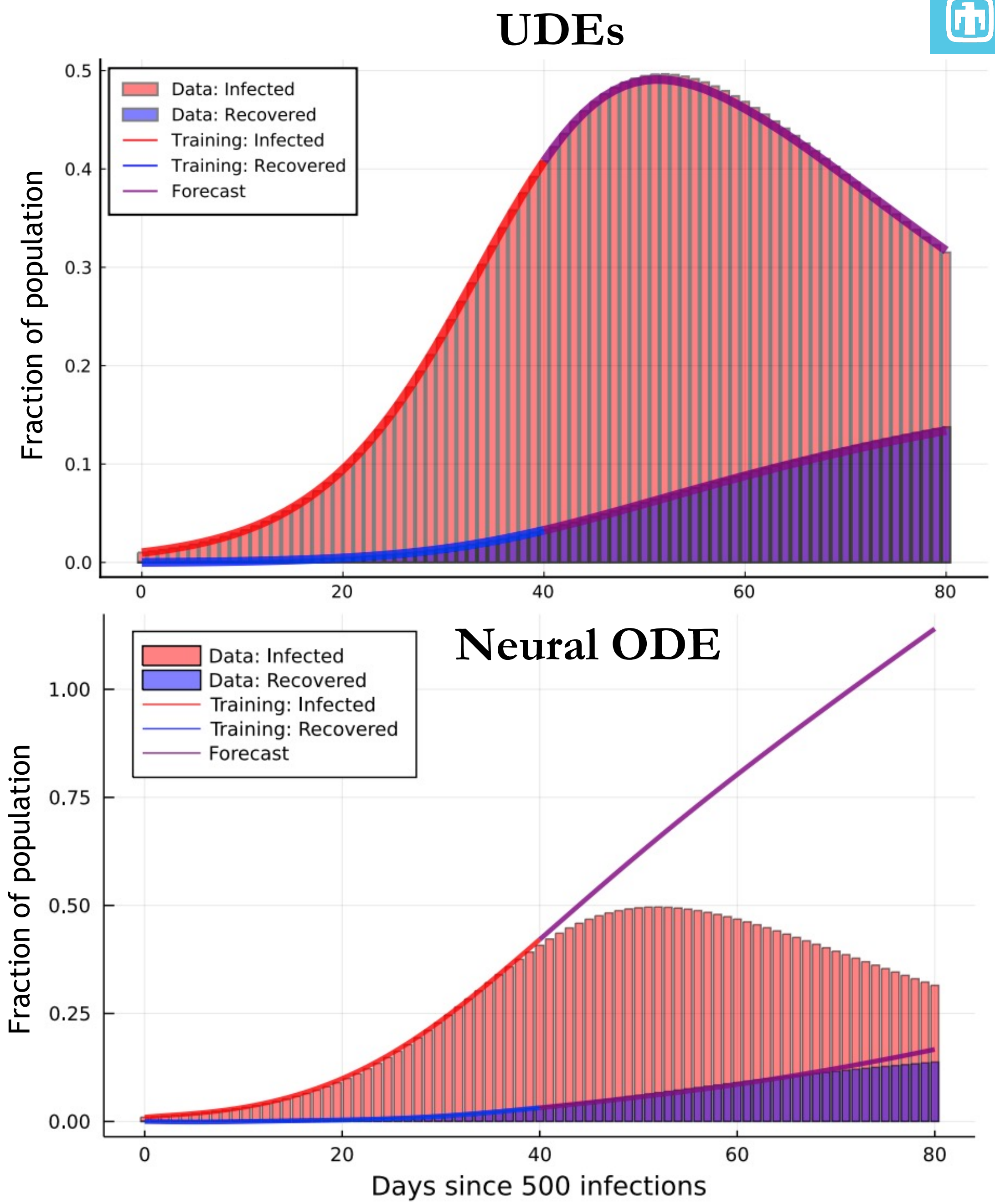
Motivation

- Universal Differential Equations (UDEs) have been successfully deployed to infer interpretable, predictive dynamics from data [1,2].
- UDEs embed ML models, e.g. neural nets (NNs) within existing scientific models:

$$\boldsymbol{u}' = F\big(\boldsymbol{u}, t, NN_\theta(\boldsymbol{u})\big)$$
$$\min_\theta \|\boldsymbol{d} - \boldsymbol{u}(\theta)\|$$

- Can be formulated to respect physical principles by construction.
- Data-efficient because making use of prior physical information.
- Can be more predictive than Neural ODEs:

$$\boldsymbol{u}' = NN_\theta(\boldsymbol{u})$$
$$\min_\theta ||\boldsymbol{d} - \boldsymbol{u}\,(\theta)||$$



Motivation

- UDEs had to date been trained with observations of all state variables.

- Used to infer isolation dynamics early in the COVID-19 pandemic, but only had access to a subset of state variables. [2]

How are inferred dynamics affected by incomplete observations, e.g. inability to observe all state variables?

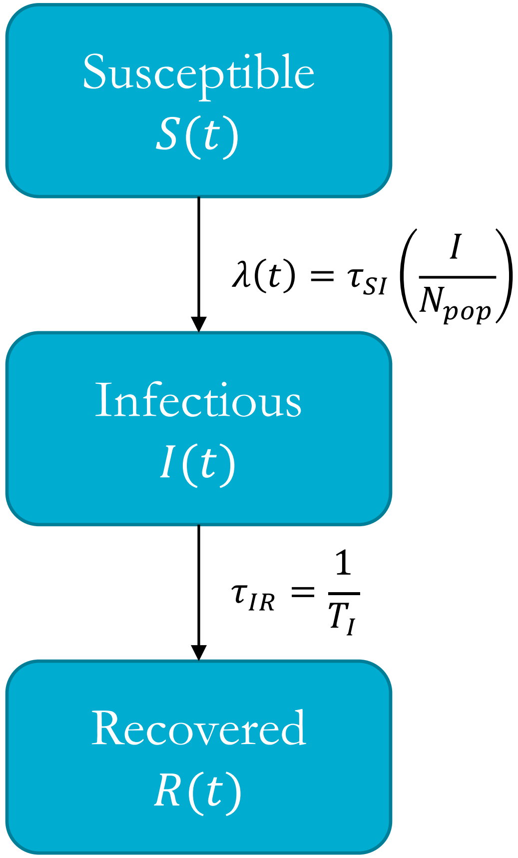
Compartment-based disease models [3]

Let $N_{pop} = S(t) + I(t) + R(t)$.

$$\frac{dS}{dt} = -\frac{\tau_{SI} I S}{N_{pop}}$$

$$\frac{dI}{dt} = \frac{\tau_{SI} I S}{N_{pop}} - \tau_{IR} I$$

$$\frac{dR}{dt} = \tau_{IR} I$$



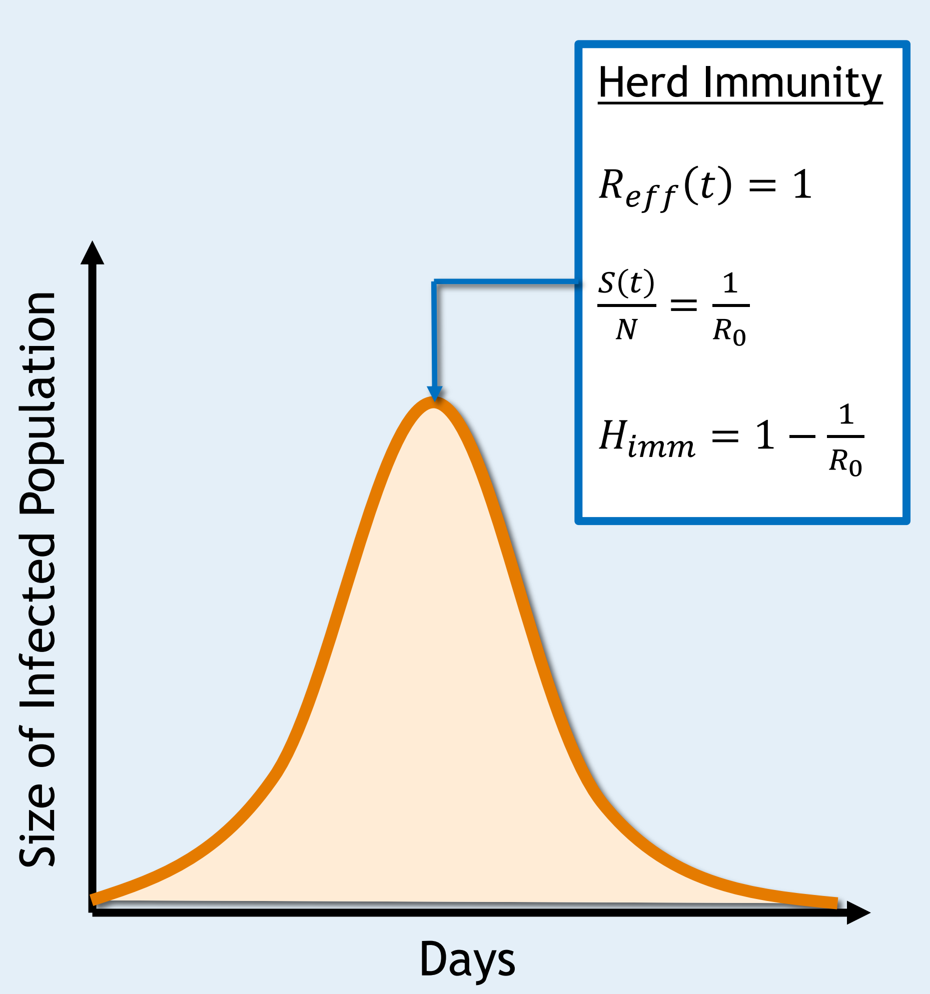
$$R_0 = \tau_{SI} T_I$$

$$R_{eff}(t) = R_0\left(\frac{S}{N_{pop}}\right)$$

- SIR a common, simple model of disease spread.
- Lots of assumptions, but can provide basic understanding of the aggressiveness of disease spread through R_0, R_{eff}.

Compartment-based disease models [3]

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



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

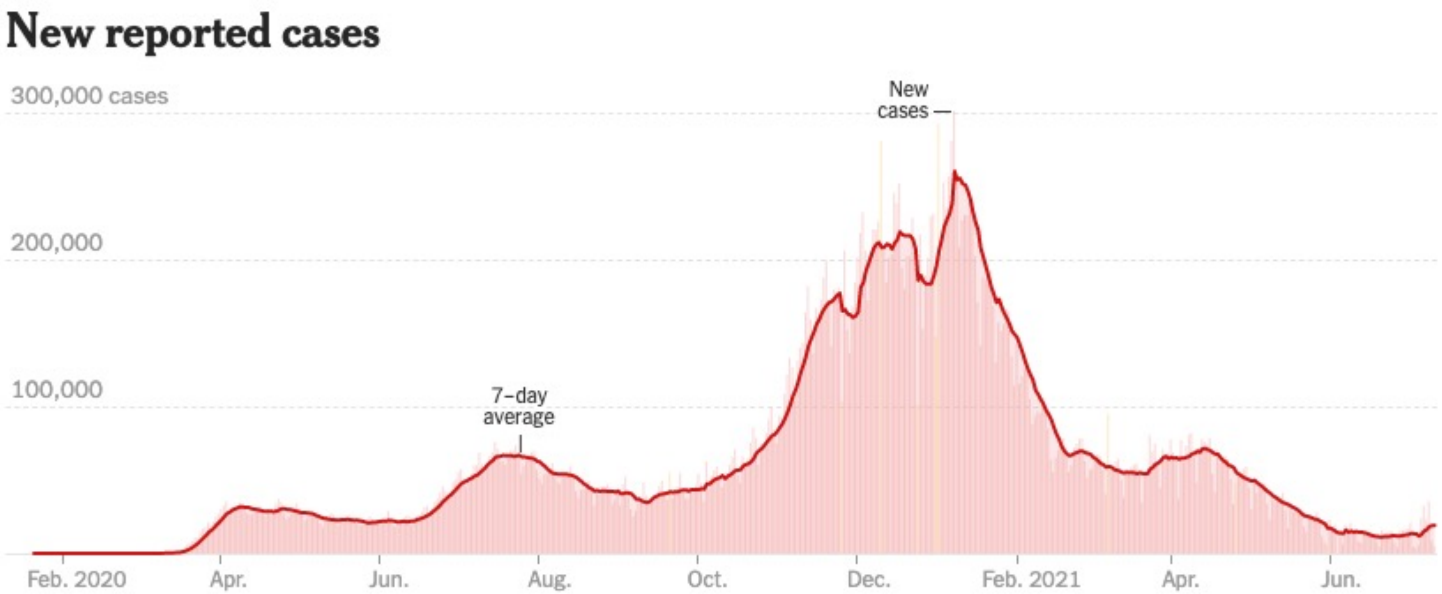


Image Credit: NYT https://www.nytimes.com/interactive/2021/us/covid-cases.html [accessed 2020/07/12]

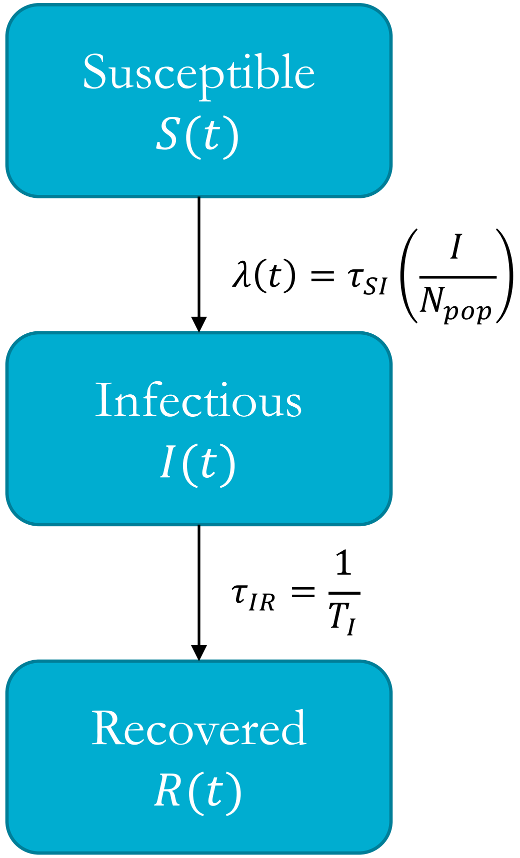
UDEs for compartment-based disease models

Let $N_{pop} = S(t) + I(t) + R(t)$.

$$\frac{dS}{dt} = -\frac{\tau_{SI} I S}{N_{pop}}$$

$$\frac{dI}{dt} = \frac{\tau_{SI} I S}{N_{pop}} - \tau_{IR} I$$

$$\frac{dR}{dt} = \tau_{IR} I$$



$$R_0 = \tau_{SI} T_I$$

$$R_{eff}(t) = R_0 \left(\frac{S}{N_{pop}}\right)$$

- Does not account for significant portion of infected population being isolated as we saw for COVID-19.
- Isolation dynamics could depend nonlinearly on all state variables.

UDEs for compartment-based disease models

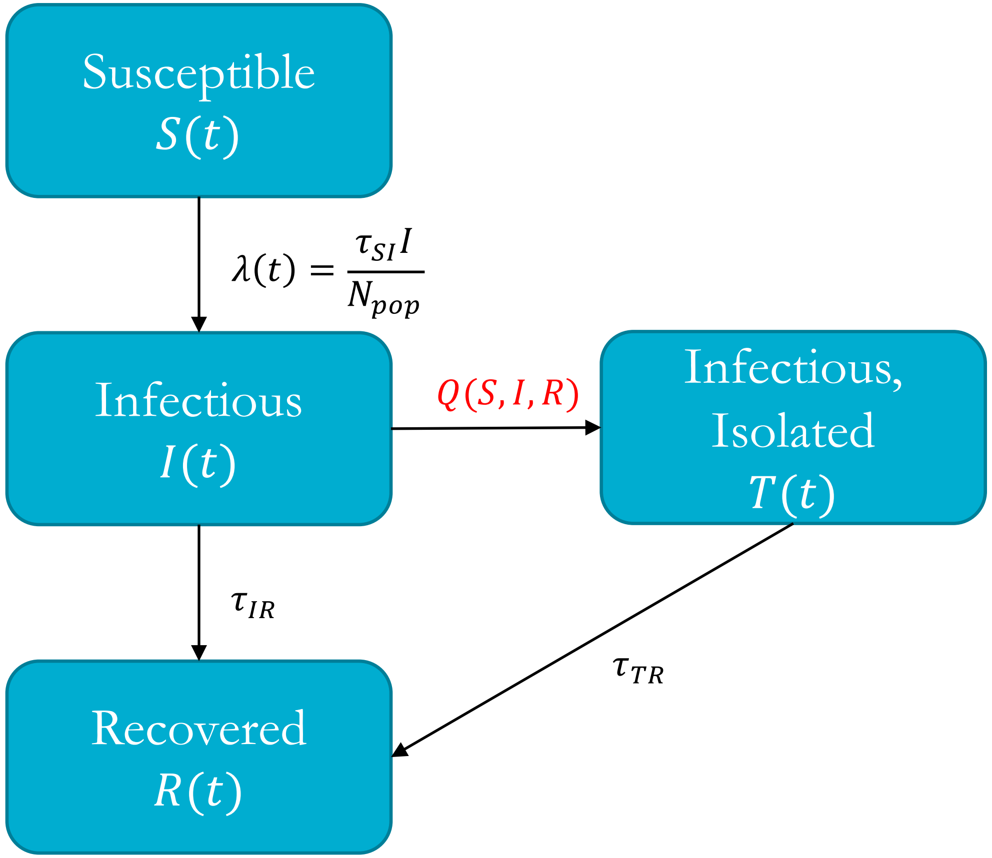
Let $N_{pop} = S + I + R + T$.

$$\frac{dS}{dt} = -\frac{\tau_{SI} IS}{N_{pop}}$$

$$\frac{dI}{dt} = \frac{\tau_{SI} IS}{N_{pop}} - \tau_{IR} I - Q(S, I, R) I$$

$$\frac{dR}{dt} = \tau_{IR} I + \tau_{TR} T$$

$$\frac{dT}{dt} = Q(S, I, R) I - \tau_{TR} T$$



- [2] introduced isolation state T, used UDE for the nonlinear, evolving transition rate into T, denoted Q.
- Q represented with small neural network depending on S, I & R, denoted NN_Q.
- By definition constrained to conserve population, i.e. $\frac{dN_{pop}}{dt} = 0$.

Inferring transition into quarantine with incomplete data

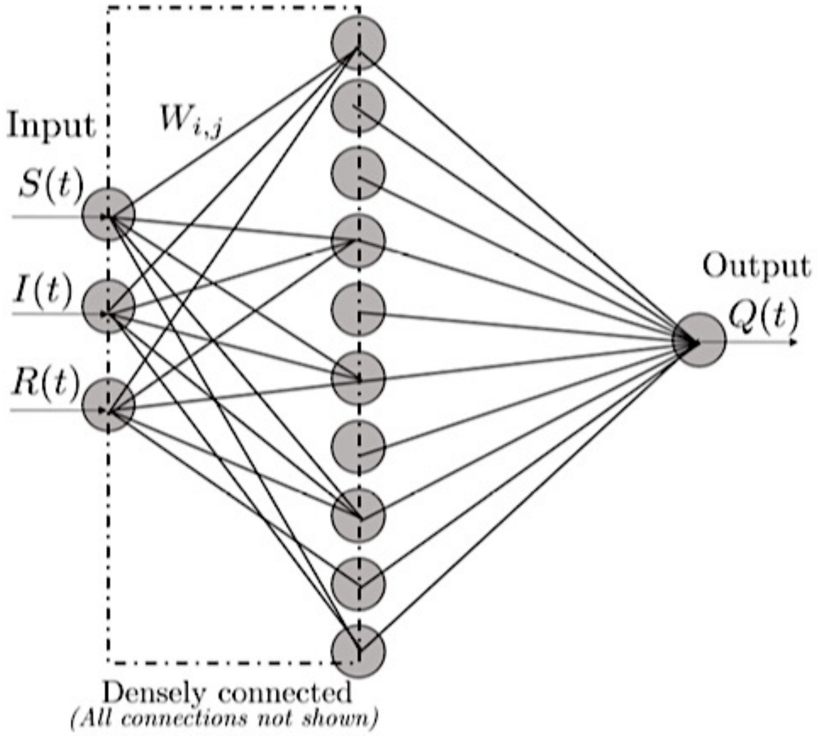
- (Dandekar 2020 [2]) used observations of I, R to infer transition rates (including Q) for COVID-19.
  - *Only a subset of the state variables could be observed.*
- How does this affect the ability to recover "useful" information about Q & disease dynamics?

Plan

- Generate synthetic data with prespecified NN_Q, τ_*.
- Infer NN_Q & transition rate parameters τ_* from combinatorial subsets of state variable observations.
  - Data = $[I, R, T], [I, R], [I, T], [R, T], [I], [R], [T]$
- Study MSE of inferred Q vs "true" Q for each dataset to determine when inference degrades.

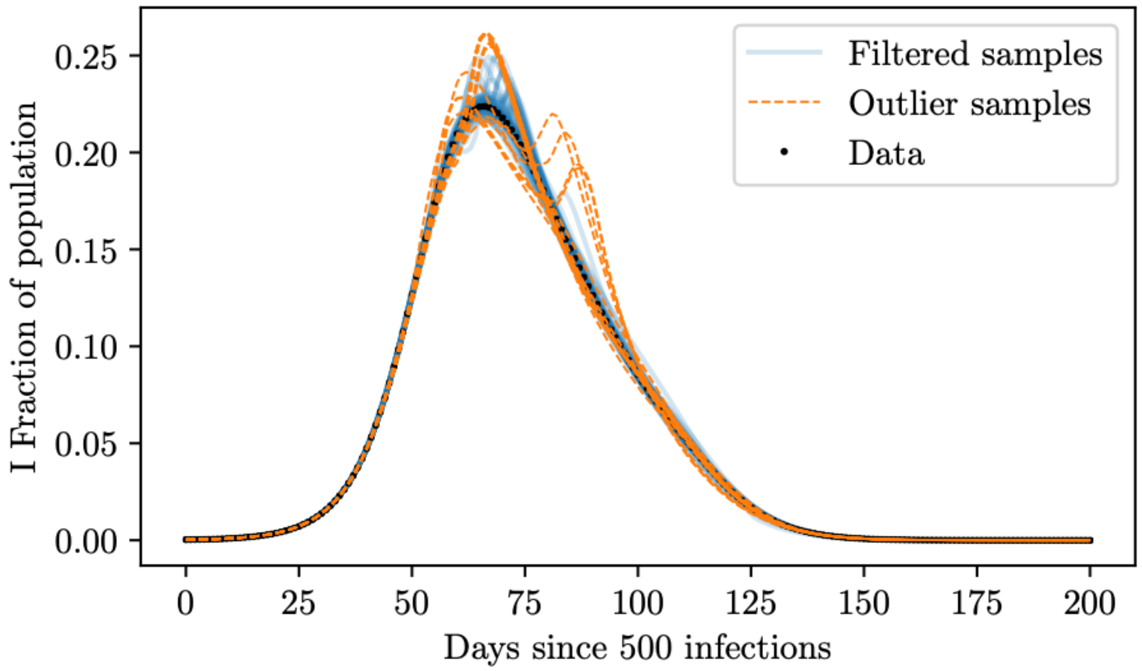
Problem Specification/State of Knowledge

- NN_Q: fully-connected NN of depth 1, width 10, ReLU activation functions.
  - Architecture the same for data generation & inference.
- Synthetic data not corrupted by noise, full time trace used.
- Initial condition assumed known.
- All transition rate parameters τ_* uncertain; distributions derived from literature.
- NN parameters and τ_* trained using ADAM (learning rate 10^{-2}, 10^5 iterations).
- Models and training implemented using Julia's SciML libraries: https://sciml.ai



Ensemble training for robust learning and uncertainty quantification

- **Challenges:**
  - Optimization can get stuck in local minima
    - Solution sensitive to initial guess
  - Uncertainty in parameters

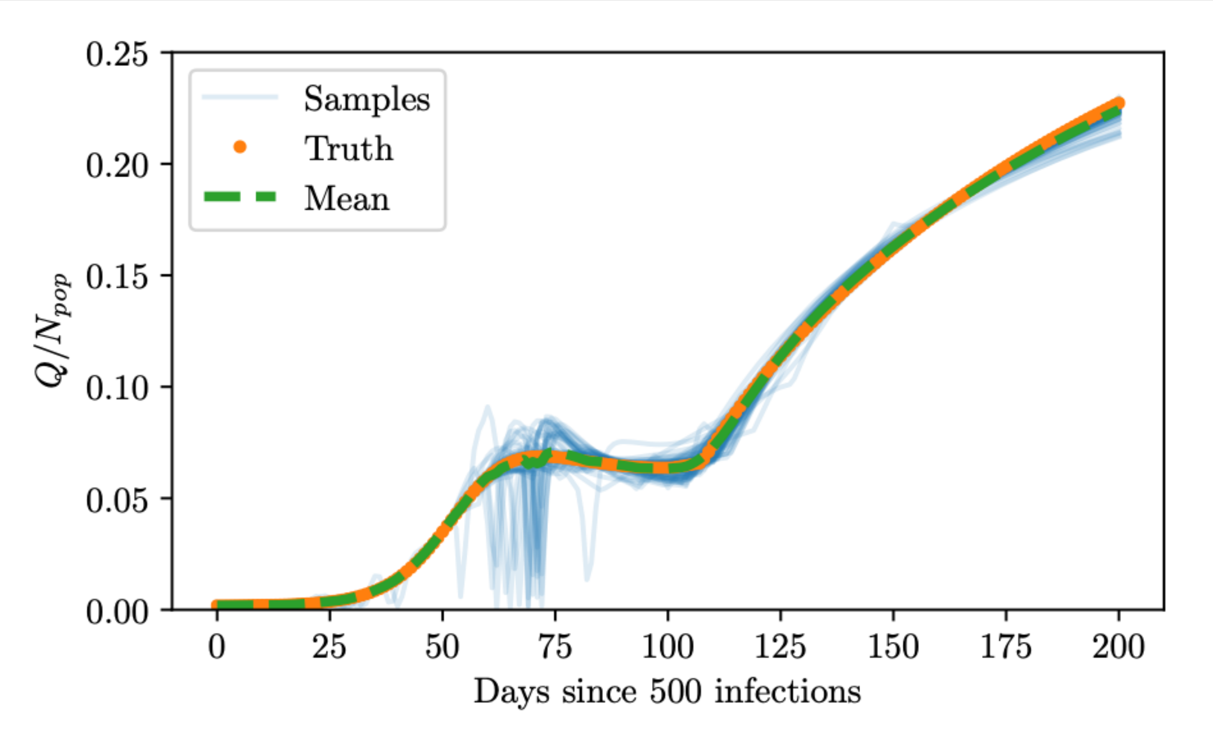


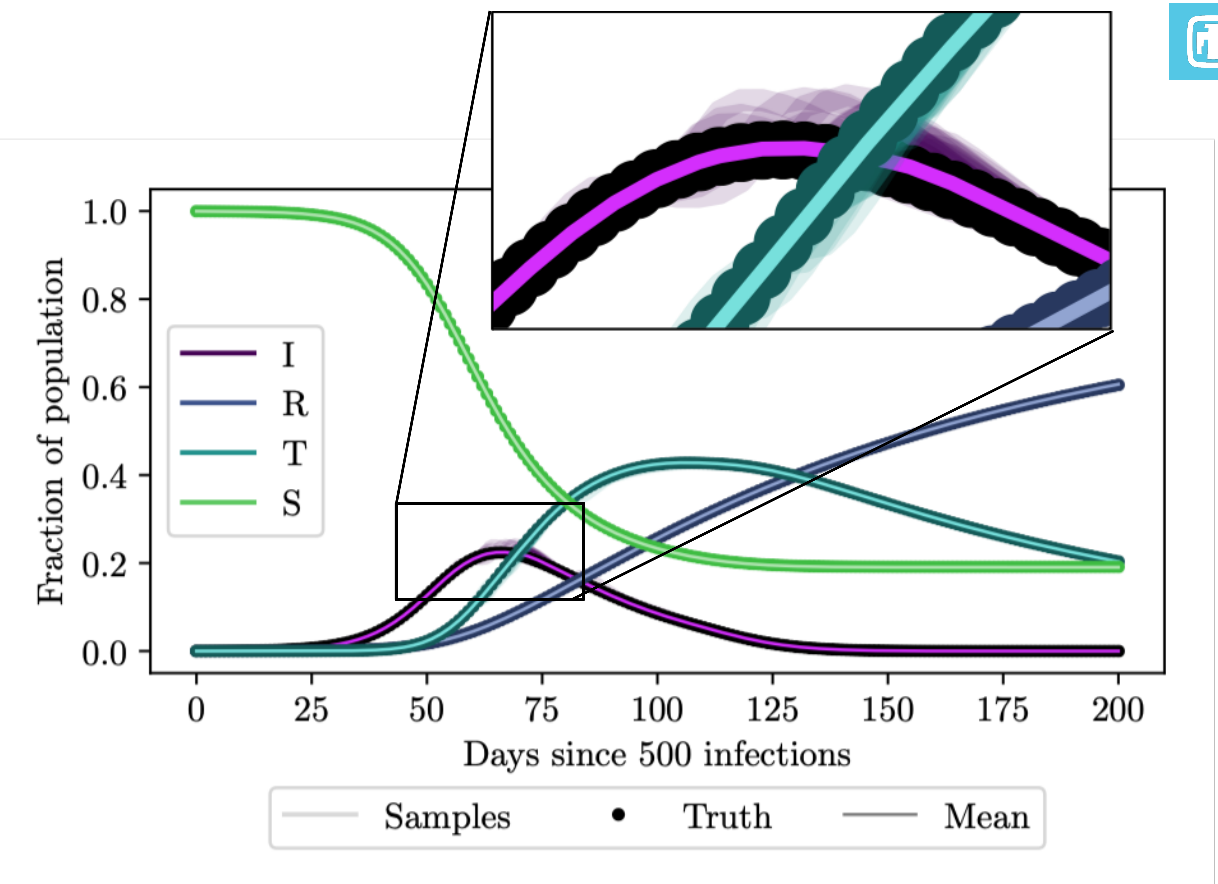
- Generated ensemble of training solutions by generating 100 random initial guesses.
  - NN parameters from Glorot initialization
  - Distributions for transition rate parameters derived from literature.

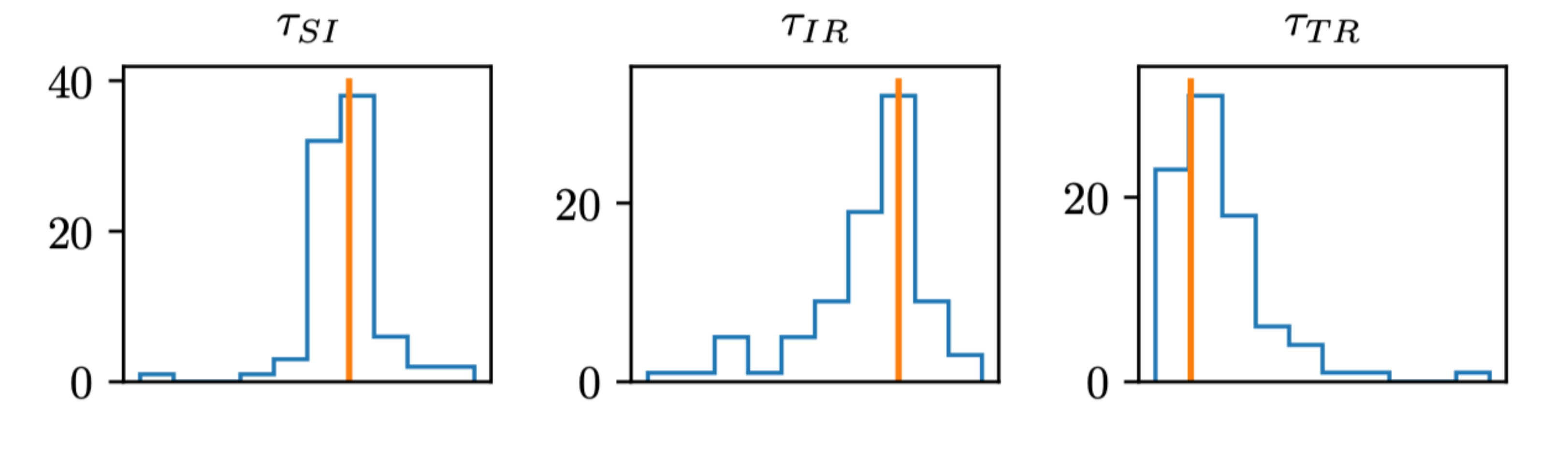
- Mitigates effect of local minima in NN training.
  - Filtered out outlier ensemble members (those with very large mean-squared error).

- Provides uncertainty information about training results. How much spread in
  - unobserved state variable trajectories?
  - optimized transition rate parameters?
  - Q trajectories?

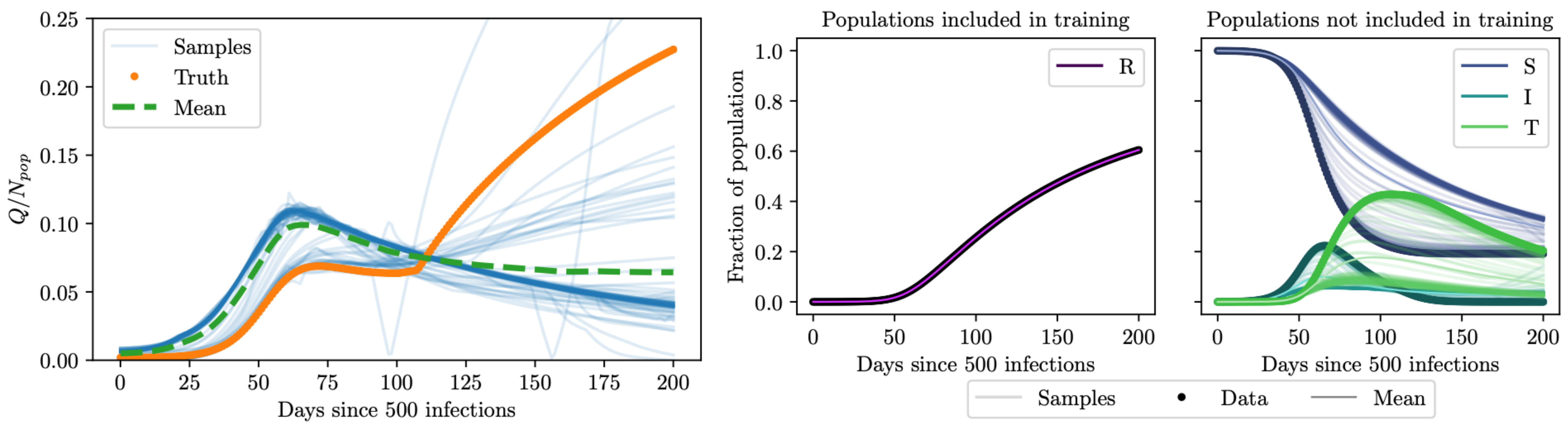
Training results, data = $[I, R, T]$

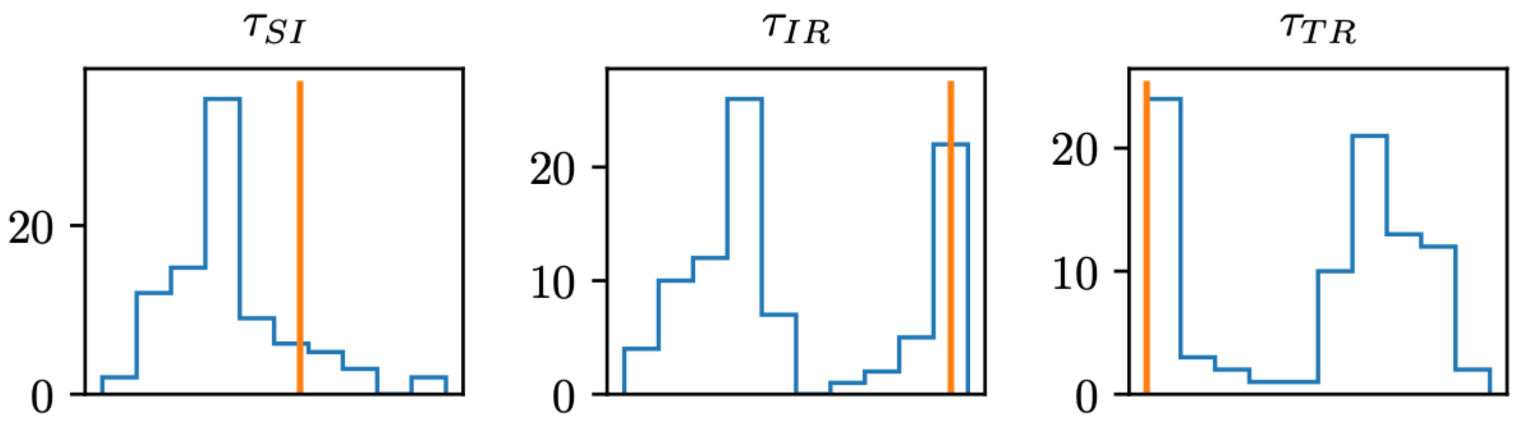






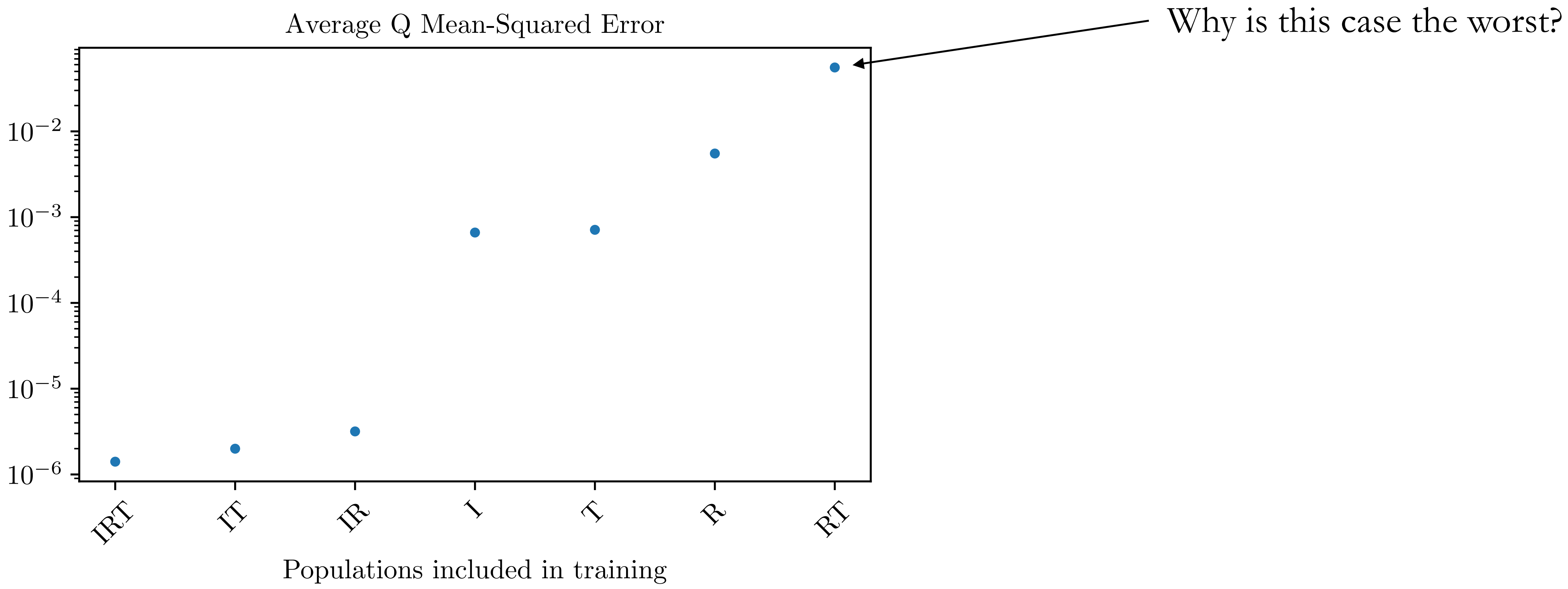
Training results, data = $[R]$



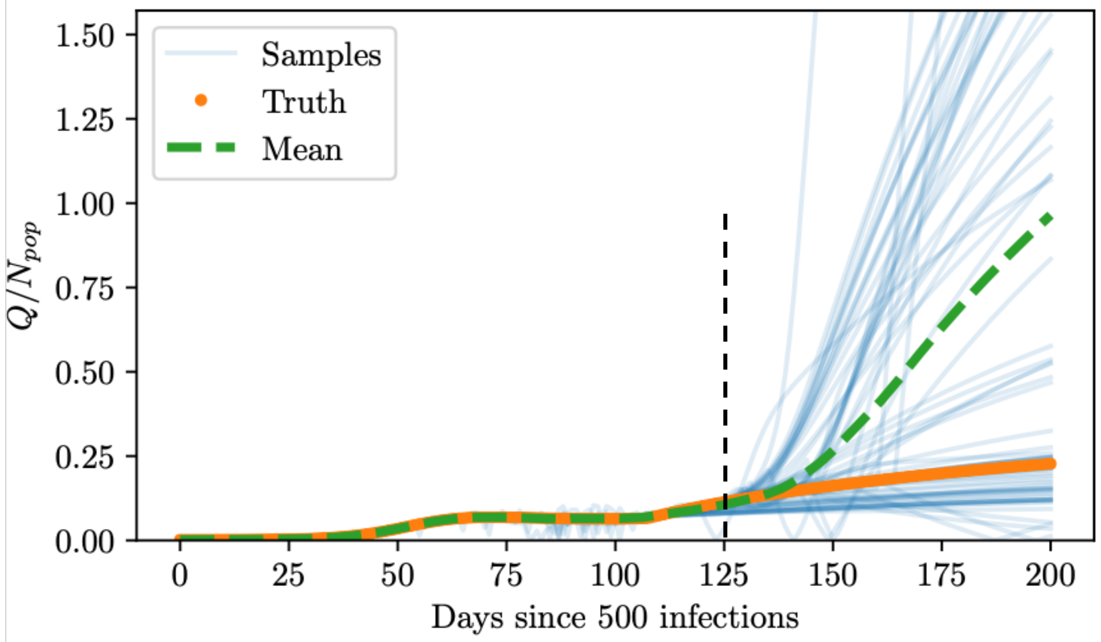


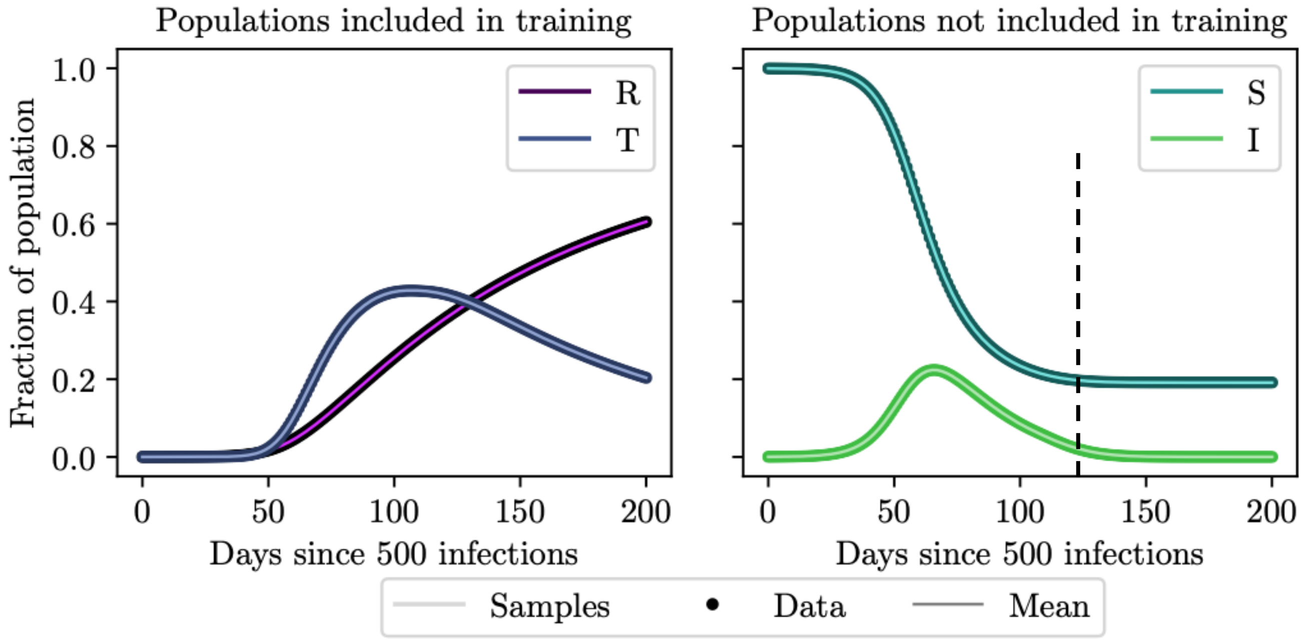
Ranking Q recovery by data subset

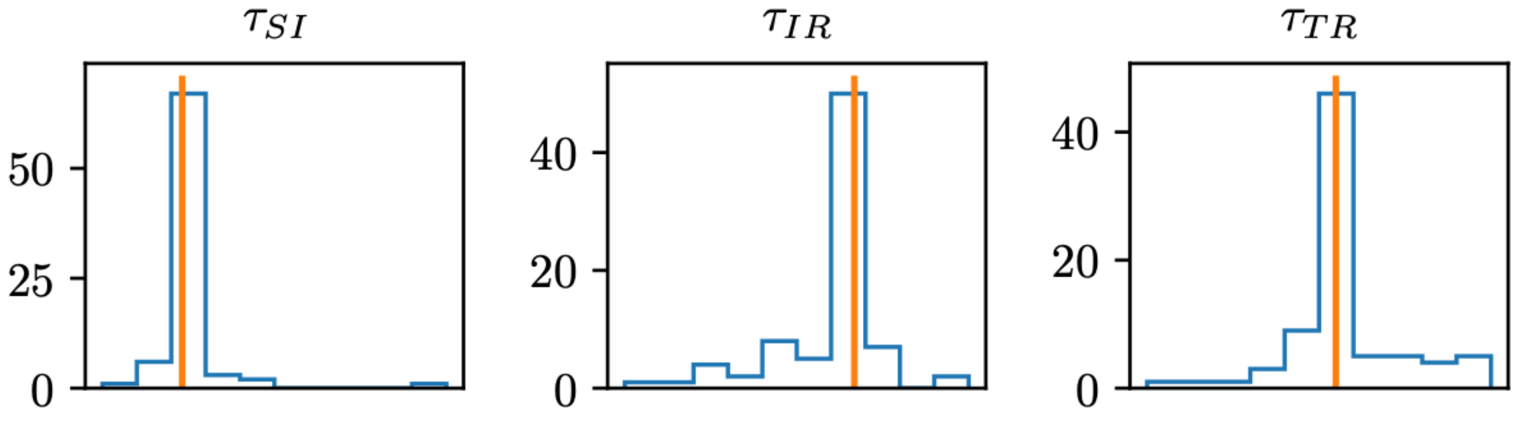
- Computed MSE of ensemble-mean (average) $\bar{Q}$ vs. true Q used to generate the data.
- Ranked data scenarios by MSE.



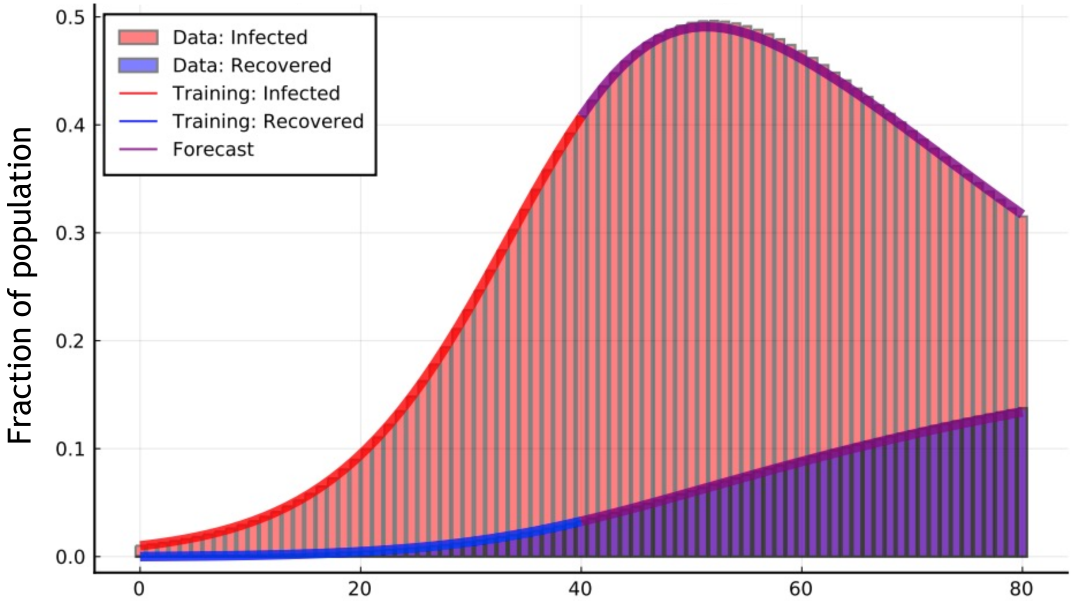
Training results, data = $[R, T]$







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: Error in observed state variables extrapolated beyond time horizon of training data.
  - Incorporate prediction uncertainty using Bayesian neural UDEs [6] or Deep Ensembles [7].

References

1. 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*.
2. 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.
3. Hethcote, H.W., 2000. The mathematics of infectious diseases. SIAM review, 42(4), pp.599-653.
4. Charles F. Manski and Francesca Molinari. Estimating the COVID-19 infection rate: Anatomy of an inference problem. Pandemic Econometrics, 220(1):181–192, January 2021.
5. Wycliffe E Wei, Zongbin Li, Calvin J Chiew, Sarah E Yong, Matthias P Toh, and Vernon J Lee. Presymptomatic transmission of SARS-CoV-2—Singapore, january 23–march 16, 2020. Mor- bidity and Mortality Weekly Report, 69(14):411, 2020. Publisher: Centers for Disease Control and Prevention.
6. Dandekar, R., Chung, K., Dixit, V., Tarek, M., Garcia-Valadez, A., Vemula, K. V., & Rackauckas, C. (2020). Bayesian neural ordinary differential equations. arXiv preprint arXiv:2012.07244.
7. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2016). Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv preprint arXiv:1612.01474.

Appendix

Initial sampling of transition rate params

- τ_{SI} bounds reported directly in [4] for several locations; used max and min over locations to define a uniform distribution.

- No direct bounds reported for τ_{IR}.

- When one pathway out of a population, transition rate is inverse of residence time in population, i.e.

$$\tau_{IR} = \frac{1}{T_I}.$$

- Instead defined distribution on T_I.

- $T_I = T_{presymptom} + T_{postsymptom}$

- $T_{presymptom} \sim \mathcal{U}[1,3]$ days [5], $T_{postsymptom} \sim \mathcal{U}[0,10]$ days (CDC guidance for symptomatic people)

- $T_T = T_{postsymptom}$ (assuming infected won't isolate until symptom development)

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.