SAND2022-10635C

JSM 2022, August 6-11, Washington, DC

SAND***

Data-Driven Model-Form Uncertainty with Bayesian Statistics and Neural Differential Equations

Teresa Portone (SNL), **Erin C.S. Acquesta (SNL)**, Raj Dandekar (MIT), Rileigh Bandy (SNL//UC Boulder), and Chris Rackauckas (MIT)

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.

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.

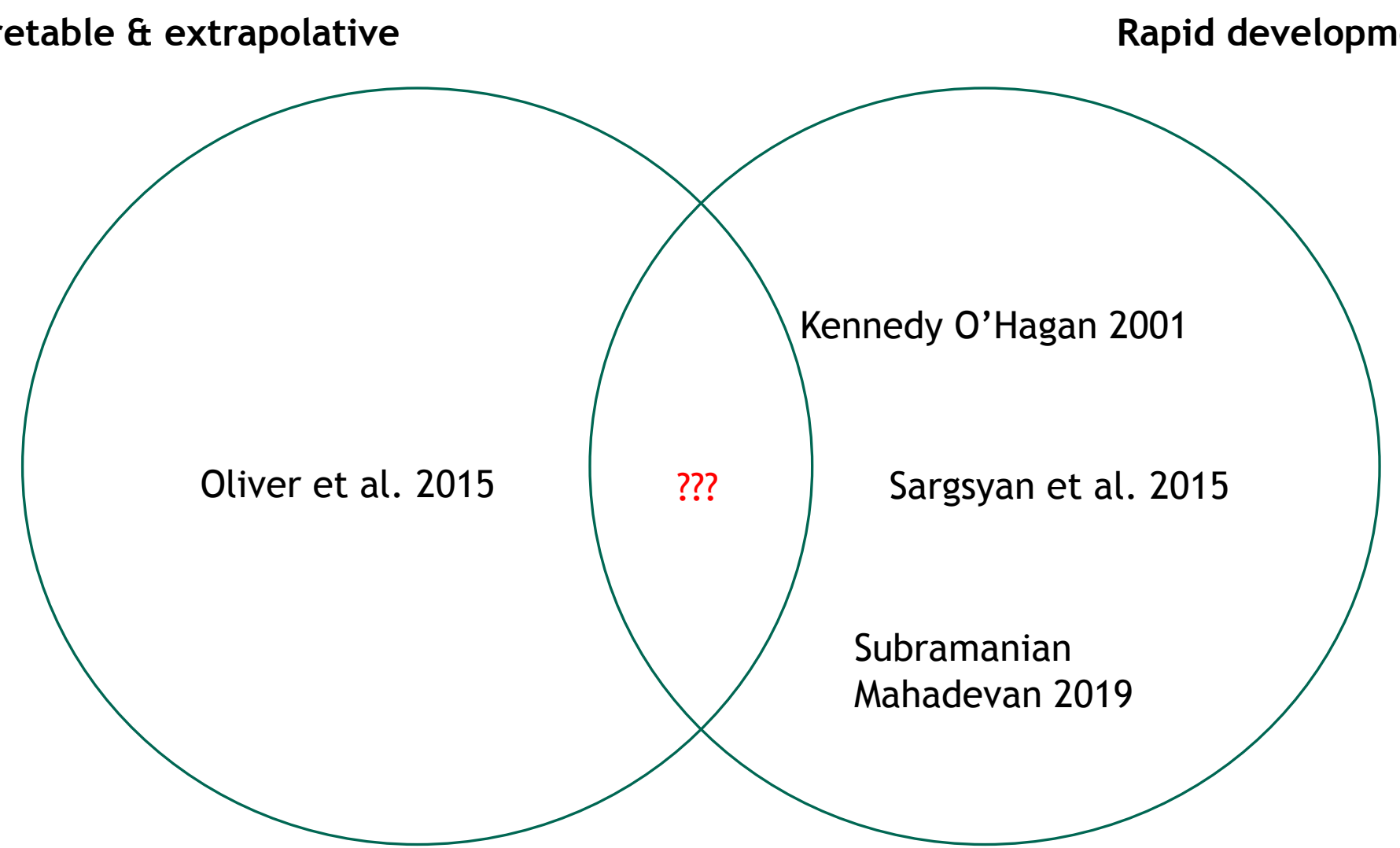
Motivation

Data-Driven Model Discrepancy

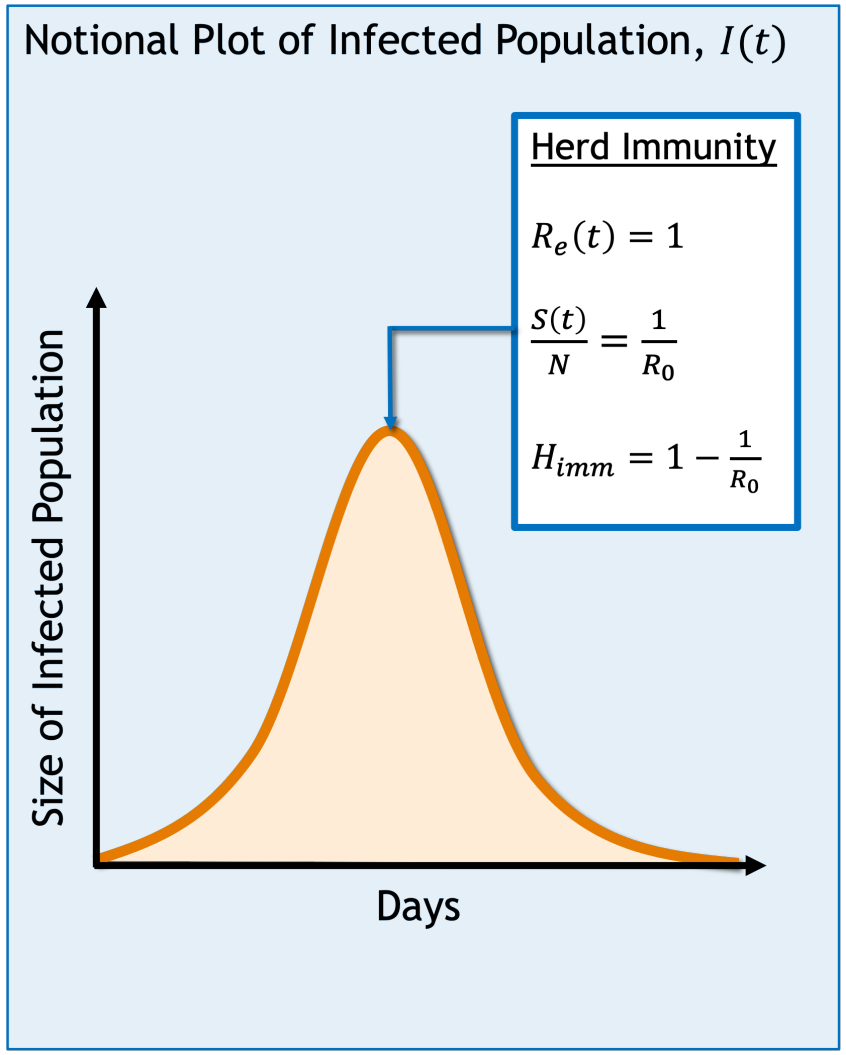
Bayesian Study

Time Varying Likelihoods

Adaptive, Interpretable, and Extrapolative Model-Form Uncertainty Motivation



Epidemiology Motivation: Model-Form Uncertainty for Predictions



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

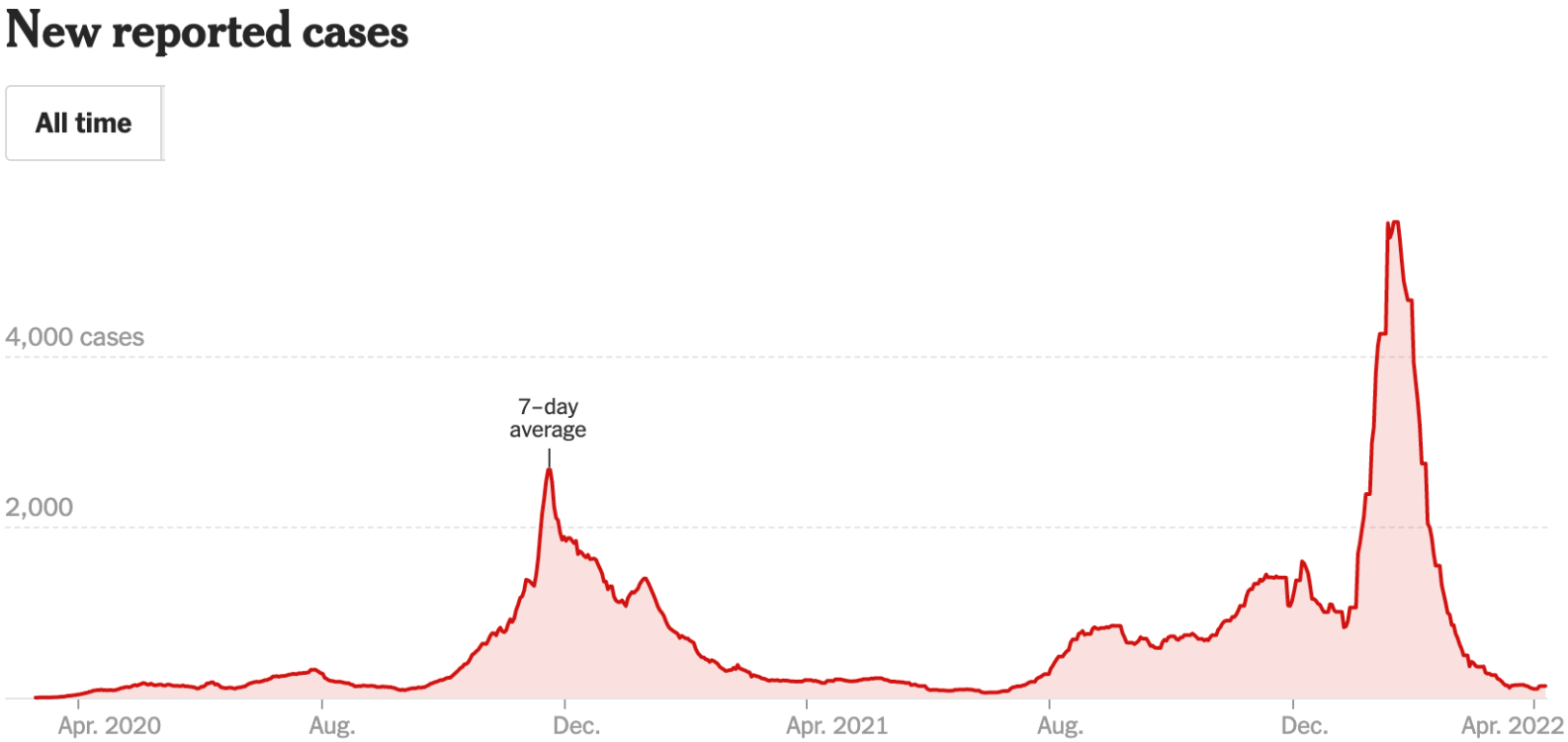


Image Credit: NYT https://www.nytimes.com/interactive/2021/us/covid-cases.html [accessed 2022/04/08]

R_0: The reproductive number, defined only at $t = 0$, when the first infectious host is presented to a population that is 100% susceptible.

$R_e(t) = R_0 \frac{S(t)}{N}$: The effective reproductive number, is the time varying rate at which new infectious cases will infect the resulting susceptible population

Motivation

Data-Driven Model Discrepancy

Bayesian Study

Time Varying Likelihoods

Universal Differential Equations (UDEs)

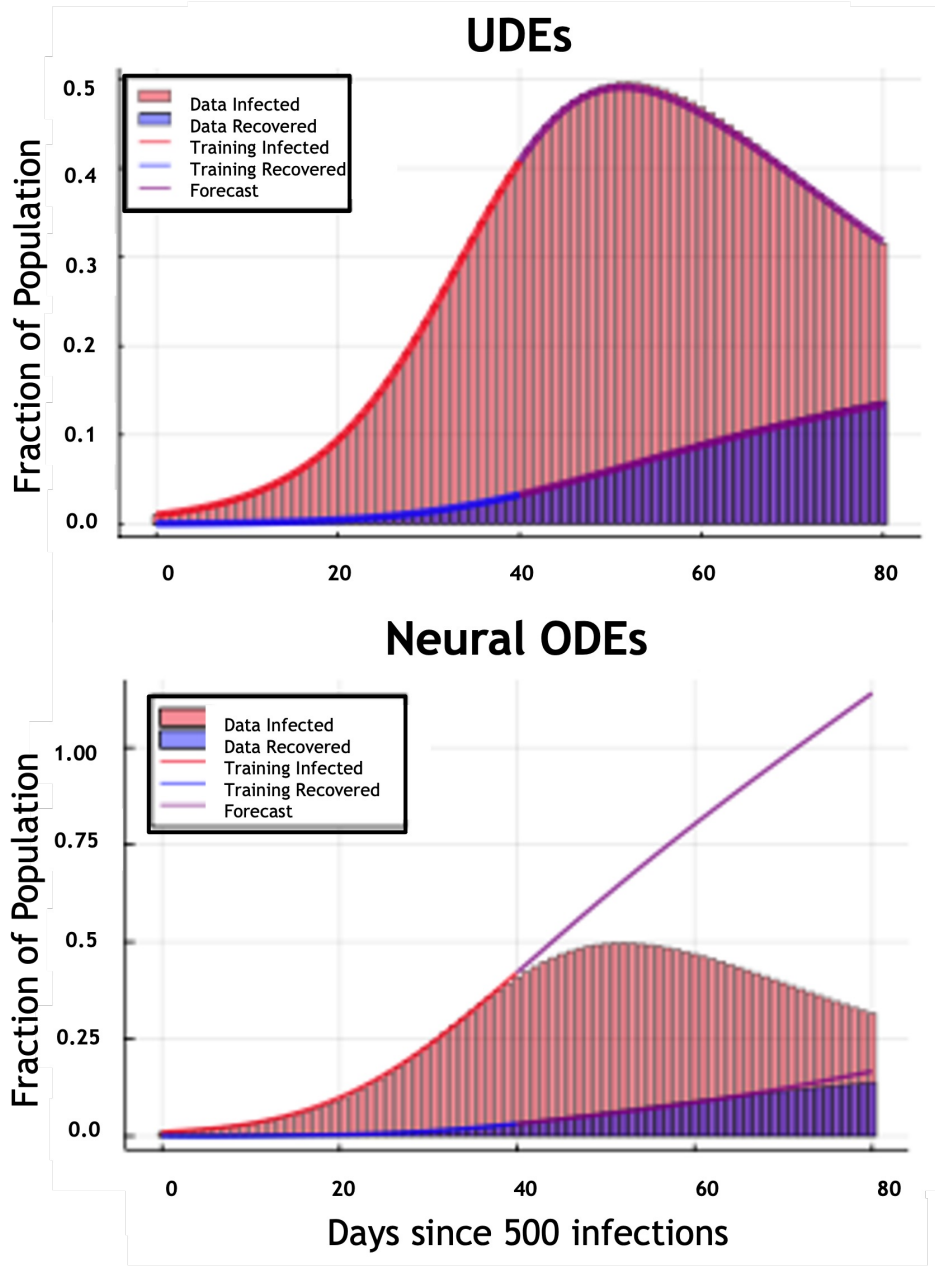
- UDEs have been successfully deployed to infer interpretable, predictive dynamics from data [[16][17]].
- UDEs embed ML models, e.g., neural networks (NNs) within existing scientific models:

$$\boldsymbol{u}' = F\left(\boldsymbol{u}, t, \theta_{ODE}, NN(\boldsymbol{u}, \theta_{NN})\right)$$
$$\min_{\theta} \|\boldsymbol{d} - \boldsymbol{u}(\theta)\|$$

  where $\theta = \{\theta_{ODE}, \theta_{NN}\}$ and $\boldsymbol{d}$ represents observation data.

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

$$\boldsymbol{u}' = NN(\boldsymbol{u}, \theta_{NN})$$
$$\min_{\theta_{NN}} \|\boldsymbol{d} - \boldsymbol{u}(\theta_{NN})\|$$



SciML: Machine Learning and Dynamical Systems

Physics-Informed Neural Networks (PINNs) [14]

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

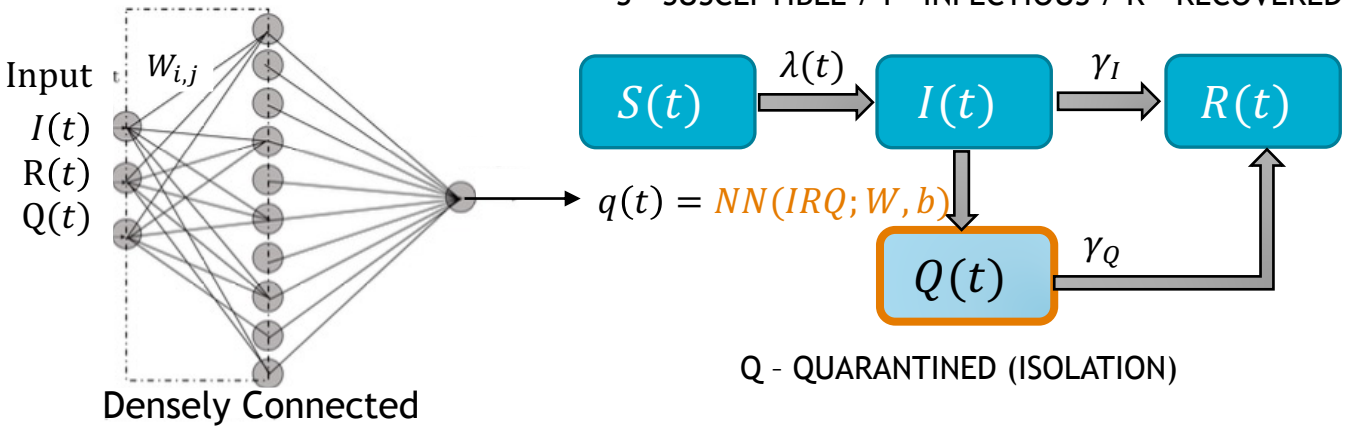
Then.... $u(t, x) = NN(W, b)$

Neural Ordinary Differential Equations (Neural ODEs) [15]

Simulating unknown dynamics for a full system of ODEs:

$$\frac{du}{dt} = NN(W, b)$$

Universal Differential Equations (UDEs) [[16][17]]



$$\frac{dS}{dt} = -\lambda(t)S(t)$$

$$\frac{dI}{dt} = \lambda(t)S(t) - \gamma_I I(t) - q(t)I(t)$$

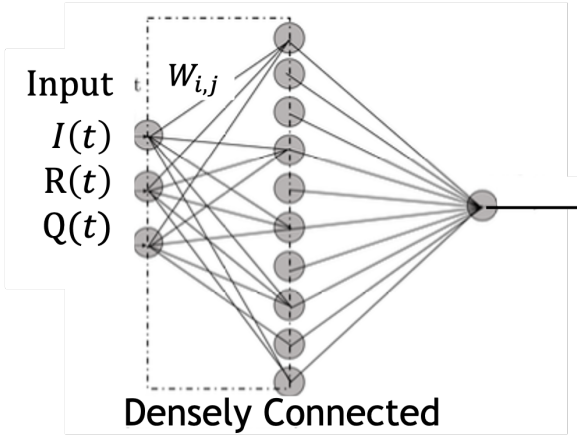
$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

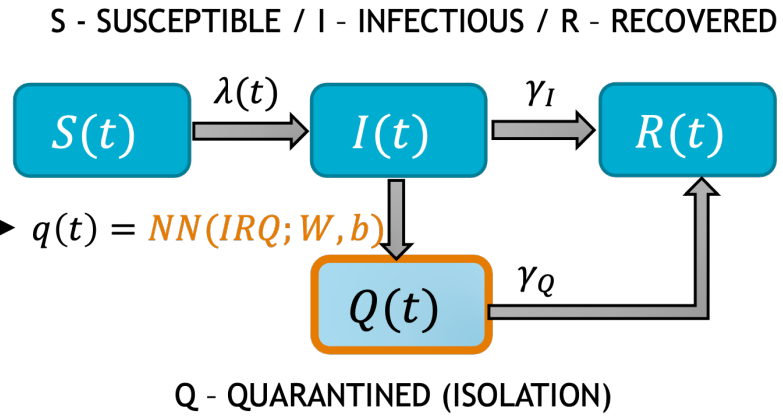
$$\frac{dQ}{dt} = q(t)I(t) - \gamma_Q Q(t)$$

Such that:

$$\lambda(t) = \beta \frac{I(t)}{S(t) + I(t) + R(t) + Q(t)}$$

UDEs for Epidemiology Compartmental Models [4]





$$\frac{dS}{dt} = -\lambda(t)S(t)$$

$$\frac{dI}{dt} = \lambda(t)S(t) - \gamma_I I(t) - q(t)I(t)$$

$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

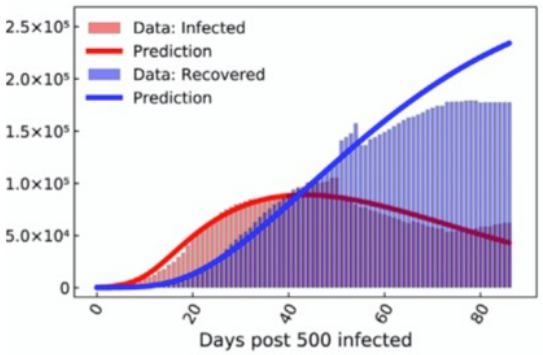
$$\frac{dQ}{dt} = q(t)I(t) - \gamma_Q Q(t)$$

Such that:

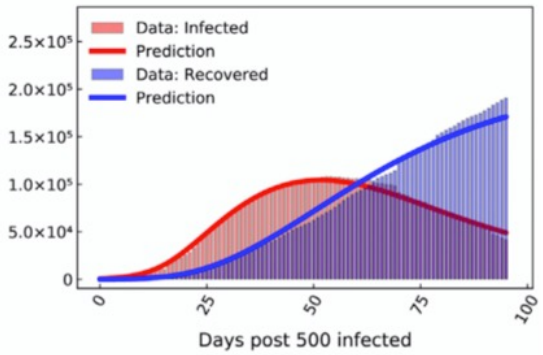
$$\lambda(t) = \beta \frac{I(t)}{S(t) + I(t) + R(t) + Q(t)}$$

Loss function:

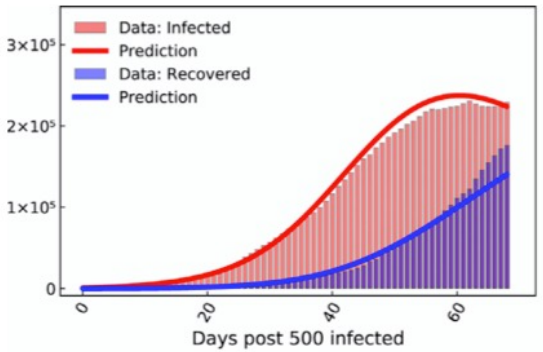
$$L_{NN}(\theta_{NN}, \beta, \gamma_I, \gamma_Q) = \left\|\log(I(t)) - \log(I_{data}(t))\right\|^2 + \left\|\log(R(t)) - \log(R_{data}(t))\right\|^2 + \left\|\log(Q(t)) - \log(Q_{data}(t))\right\|^2$$



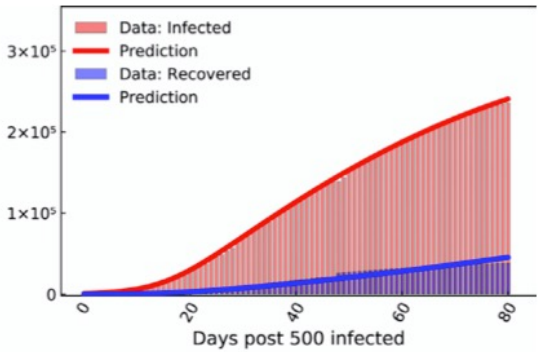
Spain



Italy



Russia



UK

Ensemble Training: Assessing Robustness and Uncertainty

Experimental Plan

1. Generate synthetic data with prespecified NN and nominal parameter values, $\Theta^{\star} = \{\Theta^{\star}_{ODE}, \Theta^{\star}_{NN}\}$
2. Learn optimal parameters $\widehat{\Theta} = \{\widehat{\Theta}_{NN}, \widehat{\Theta}_{ODE}\}$ from a subsets of observations:

   $[I, R, Q], [I, R], [I, Q], [R, Q], [I], [R], [Q]$
3. Evaluate mean-squared error (MSE) of inferred $\hat{q}(t)$ vs "true" $q^{\star}(t)$

Approach:

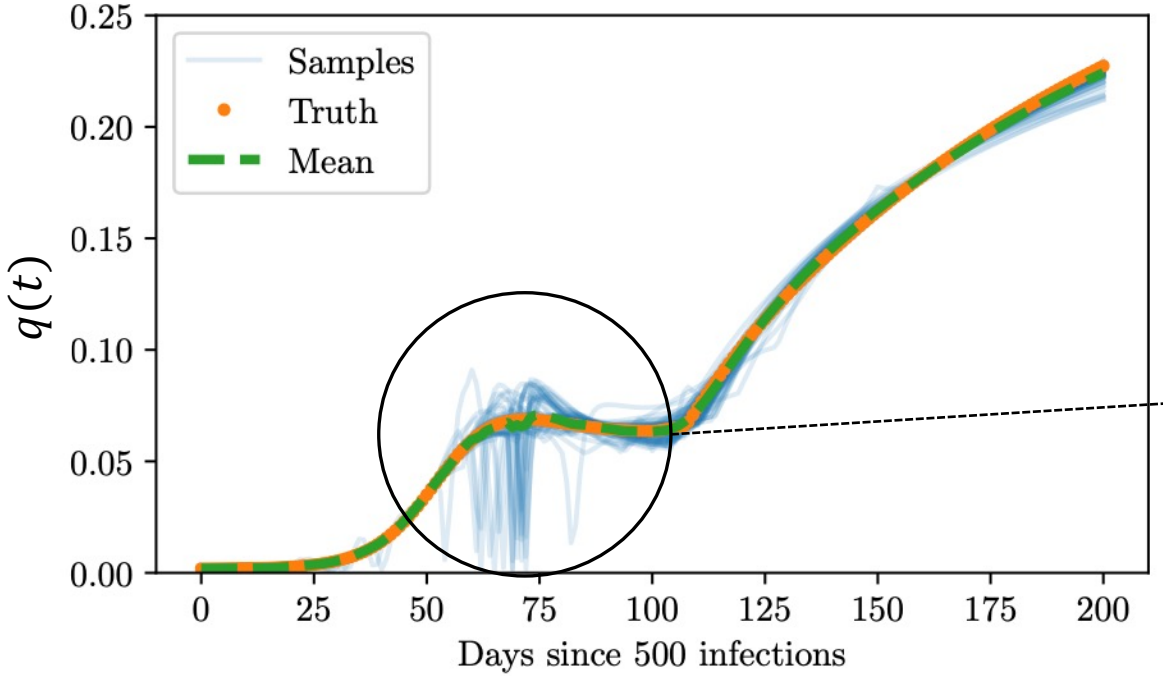
For each combination: $\{[I, R, Q], [I, R], [I, Q], [R, Q], [I], [R], [Q]\}$

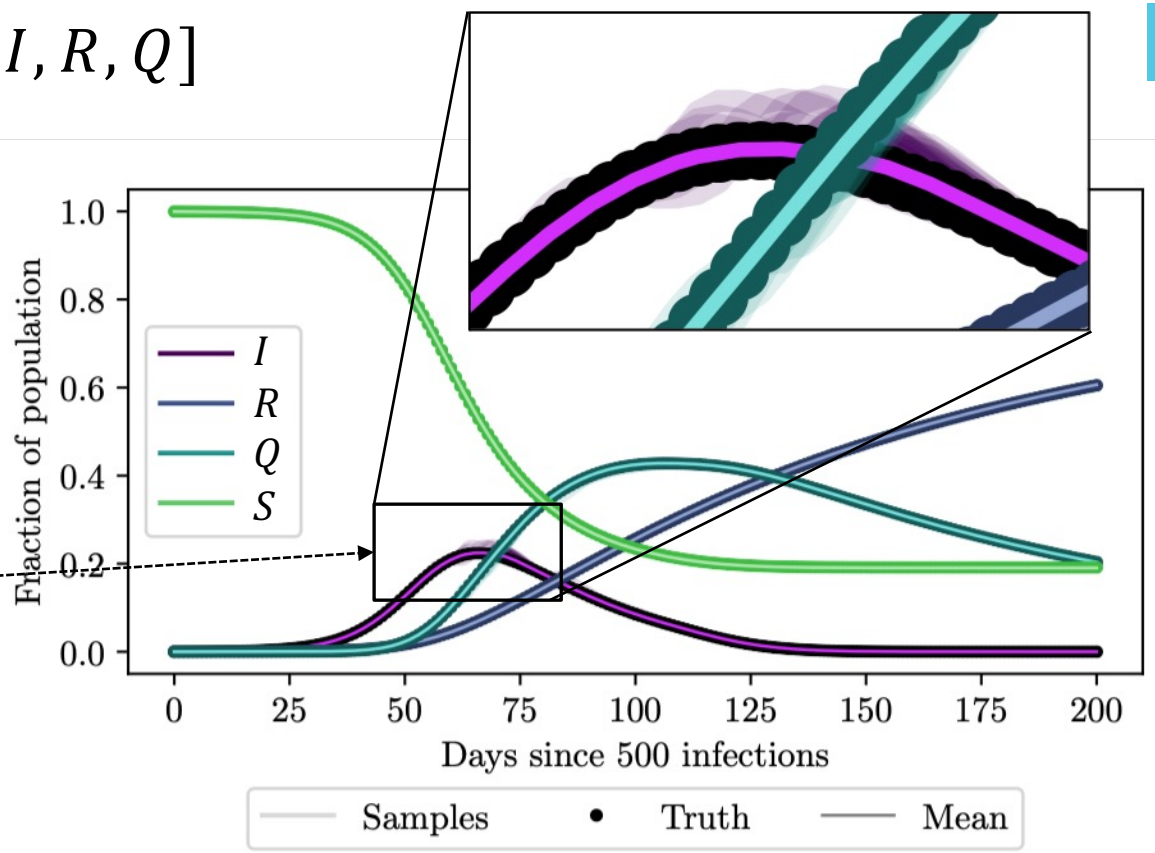
Initialize model parameters $\Theta = \{\Theta_{ODE}, \Theta_{NN}\}$

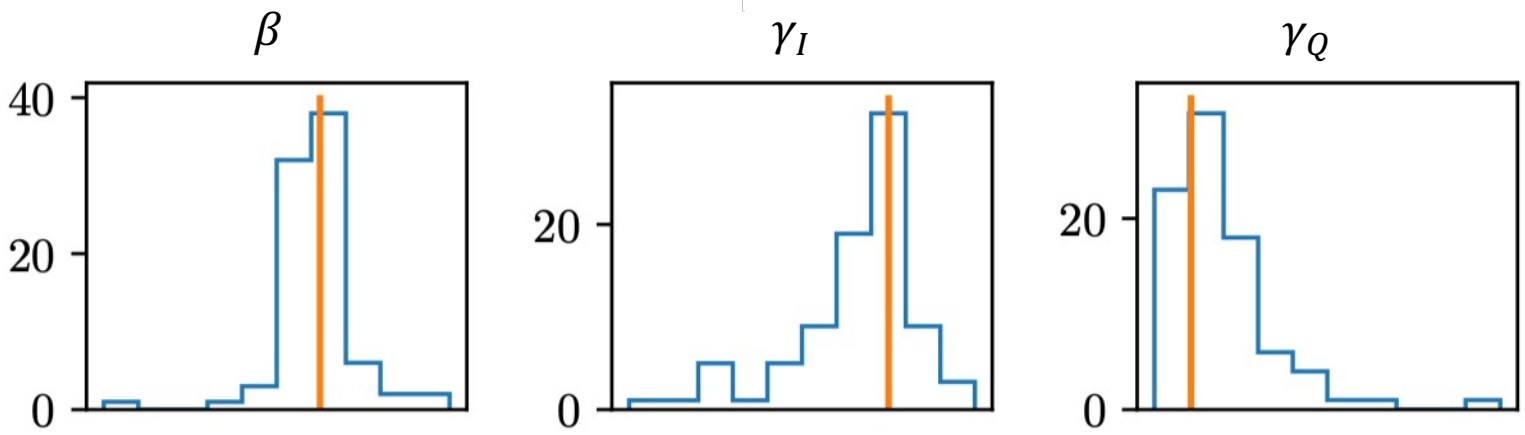
a. Θ_{ODE} sampled from distributions derived from the literature.
b. Θ_{NN} established from Glorot initialization

Run 100 training replicates to learn: $\{\widehat{\Theta}^{k}_{NN}\}$ and $\{\widehat{\Theta}^{k}_{ODE}\}$, for $k = 1, \ldots, 100$.

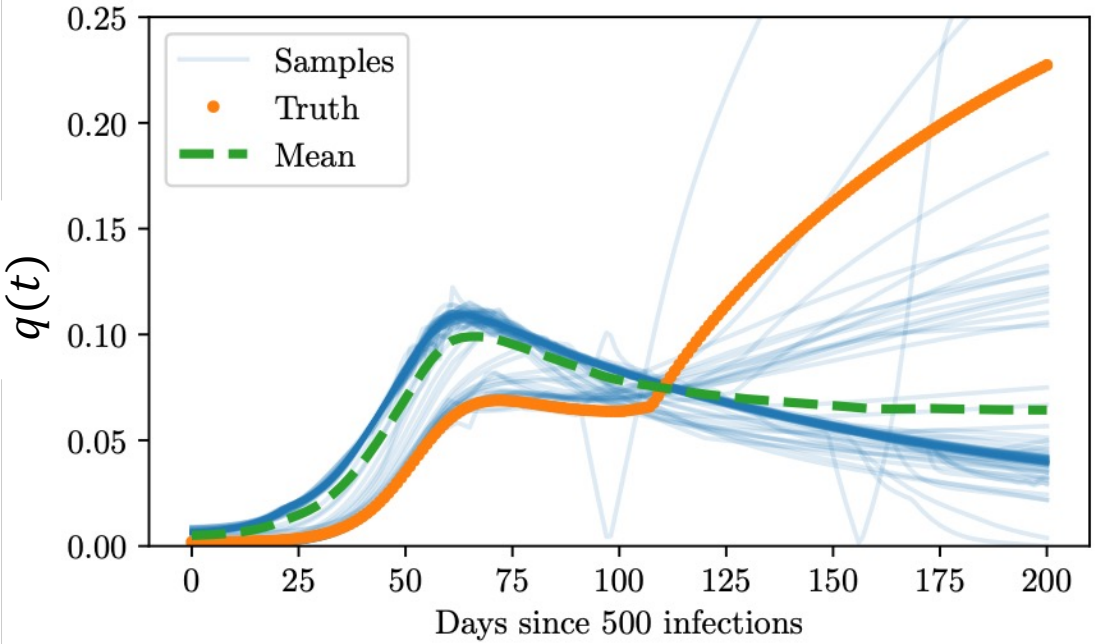
Training Results: Observable States = $[I, R, Q]$

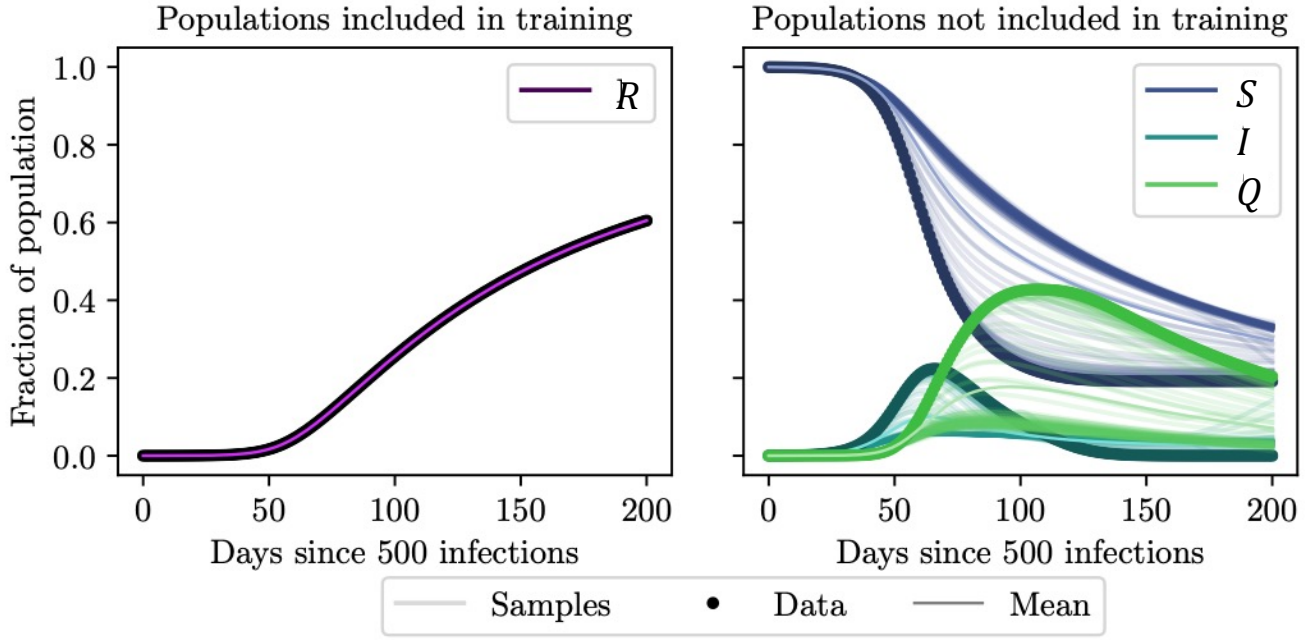


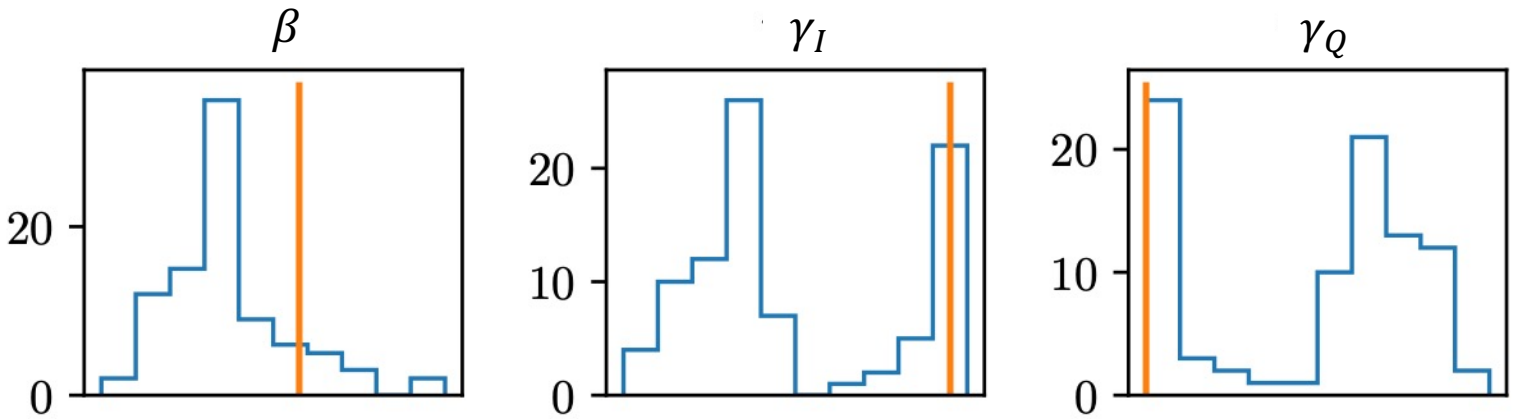




Training Results: Observable States = $[R]$







Motivation

Data-Driven Model Discrepancy

Bayesian Study

Time Varying Likelihoods

Model-Form Uncertainty: Combining UDE and Bayesian Statistics

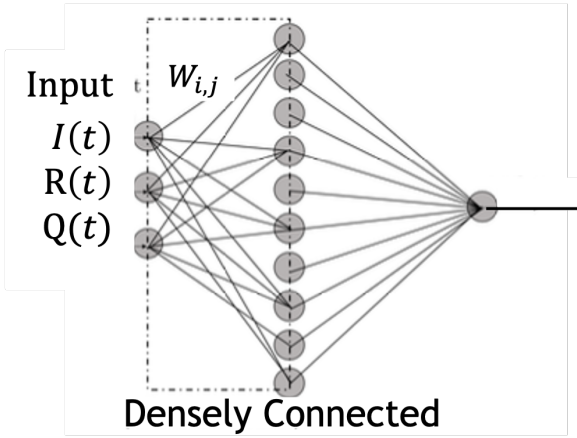
- UDEs successfully used in a deterministic setting to find "model corrections" or "missing dynamics".
- Data is NOT always informative enough to identify a single "model correction".

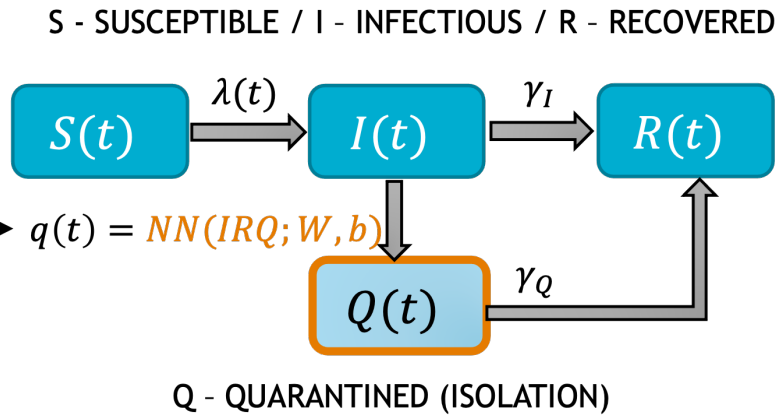
By endowing UDEs with a Bayesian parameterization, can we represent model-form uncertainty?

Challenges:

- NNs discrepancies are challenging to train even in deterministic setting.
- Bayesian methods do NOT scale well with higher dimensions

Bayesian UDE Study





$$\frac{dS}{dt} = -\lambda(t)S(t)$$

$$\frac{dI}{dt} = \lambda(t)S(t) - \gamma_I I(t) - q(t)I(t)$$

$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

$$\frac{dQ}{dt} = q(t)I(t) - \gamma_Q Q(t)$$

Such that:

$$\lambda(t) = \beta \frac{I(t)}{S(t) + I(t) + R(t) + Q(t)}$$

Loss function:

$$L_{NN}(\theta_{NN}, \beta, \gamma_I, \gamma_Q) = \left\|\log(I(t)) - \log(I_{data}(t))\right\|^2 + \left\|\log(R(t)) - \log(R_{data}(t))\right\|^2 + \left\|\log(Q(t)) - \log(Q_{data}(t))\right\|^2$$

Inferring disease parameters $[\beta, \gamma_I, \gamma_Q]$ along with NN parameters

Prior

- Disease parameters $\sim \mathcal{U}(0,2)$
- 51 NN parameters $\sim \mathcal{N}(0, (50)^2)$

Likelihood

- Synthetic data generated from SIRQ model
- Calibration data = observations of I, R, Q first 50 days
- Likelihood assumes the following error

$$d = u + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2), \qquad 2\sigma = \pm 0.1u$$

Bayesian Study Results: No Noise in Simulation Data

$$\frac{dS}{dt} = -\lambda(t)S(t)$$

$$\frac{dI}{dt} = \lambda(t)S(t) - \gamma_I I(t) - q(t)I(t)$$

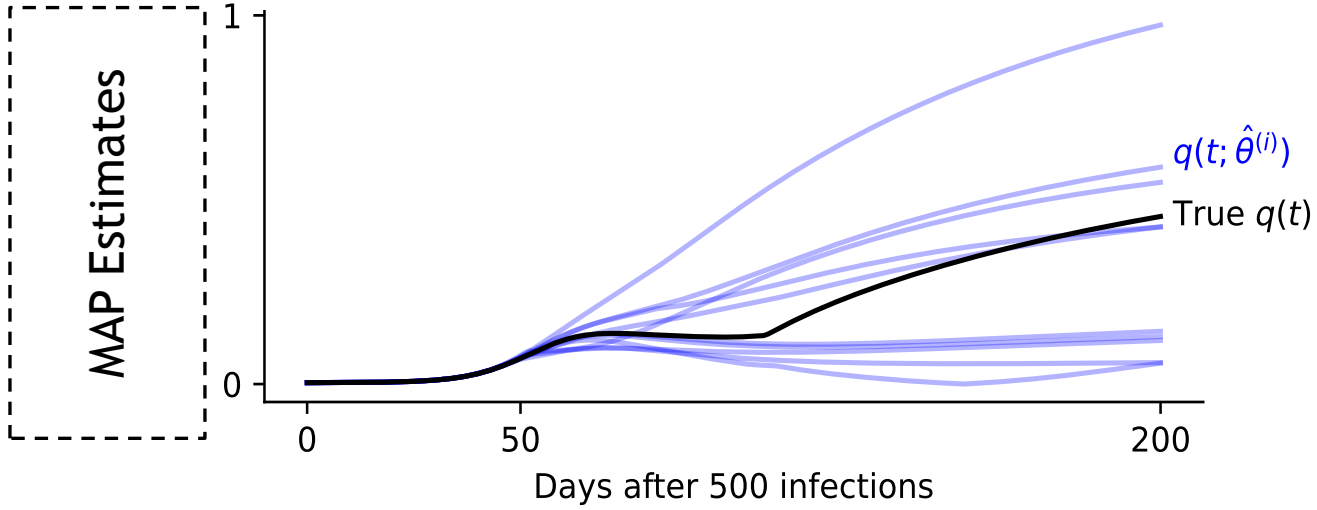
$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

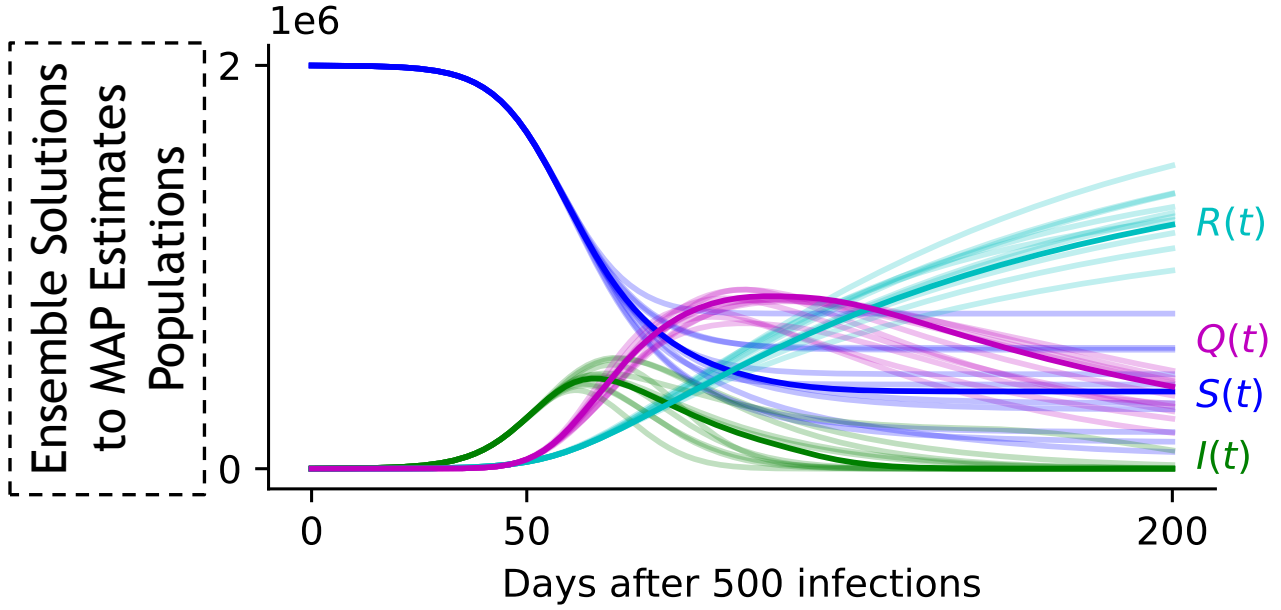
$$\frac{dQ}{dt} = q(t)I(t) - \gamma_Q Q(t)$$

Such that:

$$\lambda(t) = \beta \frac{I(t)}{S(t) + I(t) + R(t) + Q(t)}$$

$$q(t) = NN(IRQ; W, b)$$



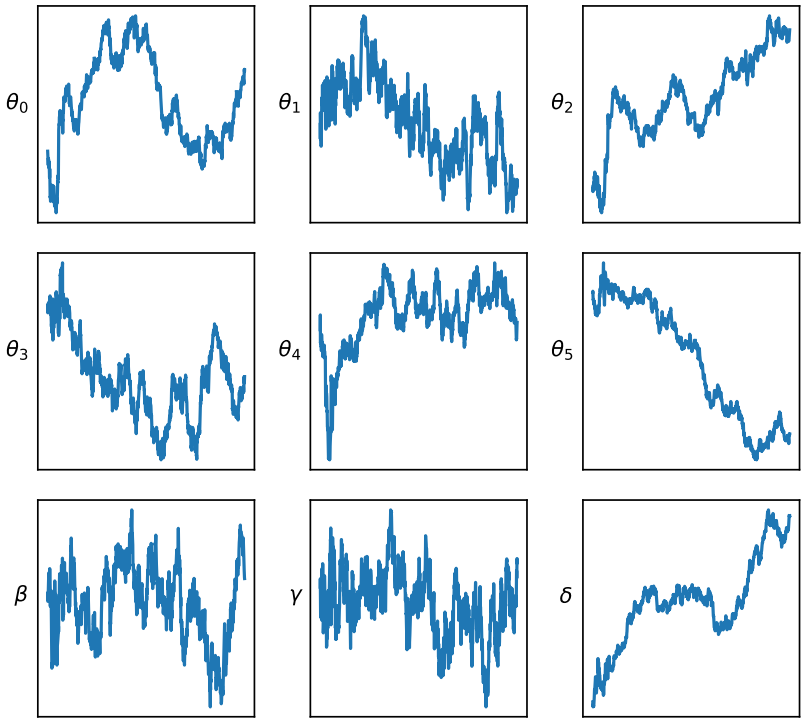


Bayesian Study Results: Indication of Complex Posterior Structure

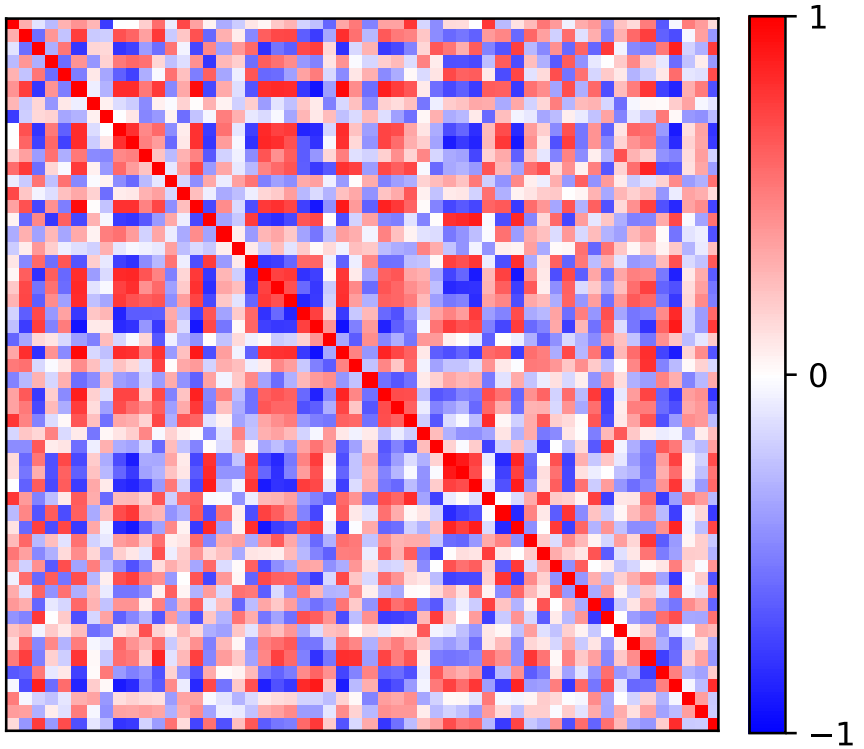
Method: NUTS, HMC variant, derivative based

- Seeded posterior approximations at MAP point

NUTS posterior chains



Correlations



2000 steps / Average acceptance rate: 0.86 / Adaptive step size

Continued and Future Research

Continued:

- Stochastic differential equations for noisy data generation.
- Time varying likelihoods as stochastic processes.
- How do neural network architectures impact validation?
- Validation metrics:
  - Mahalanobis Distance
  - Quantiles
  - The Instantaneous Reliability Metrics

Future:

- Seeded posterior approximations small perturbations away from MAP.
- Sparsity-inducing priors.
- Estimate posterior with Gaussian mixture model.

Thank You for Your Time and Attention!

For questions or follow-up discussions:

Erin Acquesta, eacques@sandia.gov

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.

References

References

Credibility

1. https://www.cmmiinstitute.com
2. https://www.sei.cmu.edu
3. Oberkampf, W.L., Pilch, M. and Trucano, T.G., 2007. *Predictive capability maturity model for computational modeling and simulation* (No. SAND2007-5948). Albuquerque, NM: Sandia National Laboratories.

ML Credibility

4. Simonyan, K., Vedaldi, A. and Zisserman, A., 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.
5. Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M. and Kim, B., 2018. Sanity checks for saliency maps. *arXiv preprint arXiv:1810.03292*.
6. Ribeiro, M.T., Singh, S. and Guestrin, C., 2016, August. " Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
7. Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (pp. 4765-4774).
8. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T., 2019, January. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*(pp. 220-229).
9. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Daumé III, H. and Crawford, K., 2018. Datasheets for datasets. *arXiv preprint arXiv:1803.09010*.

References

UQ CompSim

10. Kennedy, M.C. and O'Hagan, A., 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), pp.425-464.

Universal Approximation Theorem

11. 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.
12. Pinkus, A., 1999. Approximation theory of the MLP model. *Acta Numerica 1999: Volume 8*, *8*, pp.143-195.
13. Kratsios, A., 2019. Characterizing the Universal Approximation Property. *arXiv e-prints*, pp.arXiv-1910.
14. 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.
15. Chen, R.T., Rubanova, Y., Bettencourt, J. and Duvenaud, D., 2018. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366*.
16. 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*.
17. Dandekar, R. and Barbastathis, G., 2020. Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning. *medRxiv*.

References

Scientific Machine Learning

18. Brunton, S.L., Proctor, J.L. and Kutz, J.N., 2016. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, *113*(15), pp.3932-3937.
19. Dandekar, R., Chung, K., Dixit, V., Tarek, M., Garcia-Valadez, A., Vemula, K.V. and Rackauckas, C., 2020. Bayesian neural ordinary differential equations. *arXiv preprint arXiv:2012.07244*.

UQ for ML

20. Hüllermeier, E. and Waegeman, W., 2021. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*(3), pp.457-506.

Data Quality

21. Boyack, B.E., I. Catton, R. B. Duffey, K. R. Katsma, G. S. Lellouche, S. Levy, G. E. Wilson, and N. Zuber. "Quantifying reactor safety margins part 1: an overview of the code scaling, applicability, and uncertainty evaluation methodology." *Nuclear Engineering and Design* 119, no. 1 (1990): 1-16
22. Wilson, Gary E., and Brent E. Boyack. "The role of the PIRT process in experiments, code development and code applications associated with reactor safety analysis". *Nuclear Engineering and Design* 186, no. 1-2 (1998): 23-37
23. D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., Chen, C., Deaton, J., Eisenstein, J., Hoffman, M.D. and Hormozdiari, F., 2020. Underspecification presents challenges for credibility in modern machine learning. *arXiv preprint arXiv:2011.03395*.

Backup Slides

KOH Study

Motivation

- Universal Differential Equations (UDEs) a novel model paradigm combining classical systems of differential equations with a data-driven discrepancy term. [Rackauckas2020,Dandekar2020]
- UDEs provide deterministic calibration of the differential equations parameters while simultaneously learning a nonlinear discrepancy term at the source of discrepancy.
- Bayesian neural networks have been integrated with UDEs to provide uncertainty for the solution of the state trajectories with respect to the uncertainty in the weights and the biases of the neural network. [Dandekar2021]
- **Active Research**: Extend the methods for deterministic calibration accounting for model discrepancy at the source with Bayesian statistics to develop Bayesian calibration for situations where we can break open the black box and learn the model discrepancy at the source.

Kennedy and O'Hagan

- 20 year old state-of-the-art in Bayesian Calibration that accounts for a discrepancy term, when the computer model is treated as a black box.
- Even Kennedy and O'Hagan state in their original paper: if you can open the black box and put the discrepancy term directly at the source you can arguably do better, but their method should be more generalizable [Kennedy2001].
    - Opportunity when we open the black box: methods that do so, may help with the identifiability that the KOH framework has between parameters and model discrepancy.

What does "Better" mean? How do we compare the methods? Especially for systems of differential equations, when we put the discrepancy term at the source we change the behavior of the mathematical model.

- Larger state space
- More parameters
- Different bifurcation and phase portraits
- **Parameter estimation for one model form ≠ Parameter estimation for another model form**

Outline

- Classic Compartmental Models for Infectious Disease Transmission
  - Classic notation
  - Interpretation
  - Mathematical representations for disease phenomena
- UDEs for learning discrepancy term for Infectious Disease Transmission, accounting for the effects of quarantine.
  - QSIR [Quarantine-Susceptible-Infectious-Recovered(Removed)]
  - $q(t)$: a nonlinear transmission rate for which the Infectious population transitions to Quarantine.
- The effective reproductive number is compared between SIR and QSIR
  - $R_e(t) = R_0 \frac{S(t)}{N}$
- Notional numerical example
  - Synthetic data is generated with a known $q(t) = NN(QSIR, \Theta_{NN} \coloneqq (W, b))$
  - Kennedy and O'Hagan Bayesian calibration is applied to the SIR model that will account for the discrepancy in the solution space with a Gaussian Process (GP).
  - Examples are provided that illustrate the challenge of learning the ground truth parameter estimations because the discrepancy term is a nonlinear function in differential space.

Classic Compartmental Models: R_0 and Phase Portraits

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

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

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

$$N = S(t) + I(t) + R(t), \qquad \frac{dN}{dt} = 0$$

Phase Portraits remind us that the trajectories are not only determined by the nominal parameter values,
but also the initial value of the states.

This will be further emphasized in the notional numerical examples.

The bifurcation of the dynamical system determines the model R_0

$$R_0 = \beta\left(\frac{1}{\gamma}\right), \qquad R_e(t) = \beta\left(\frac{1}{\gamma}\right)\frac{S(t)}{N}$$

Then $R_e(t)$ determines the time varying rate at which the disease spreads as a function of the proportion of the population that is susceptible at that time.

Phase Portraits of SIR when $R_0 > 1$ and $R_0 < 1$

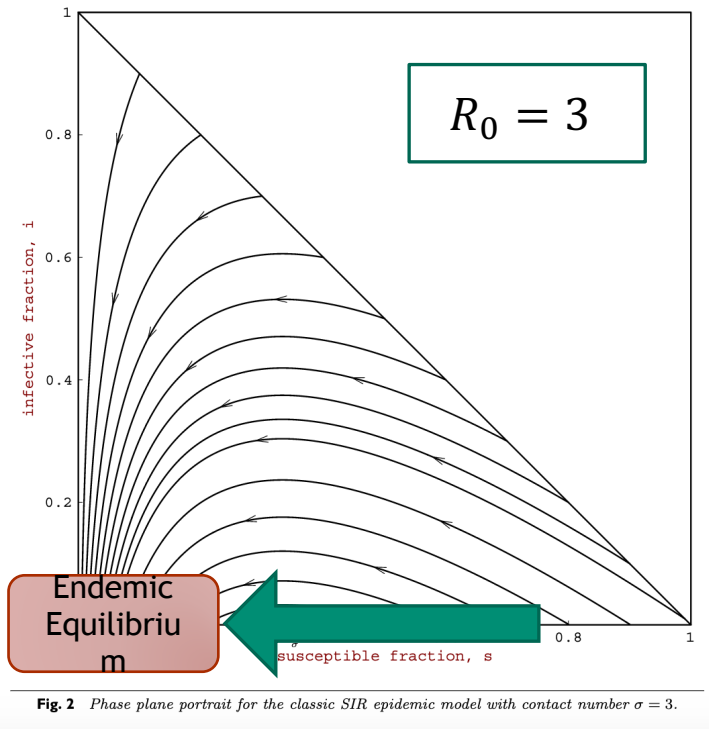


Fig. 2 *Phase plane portrait for the classic SIR epidemic model with contact number $\sigma = 3$.*

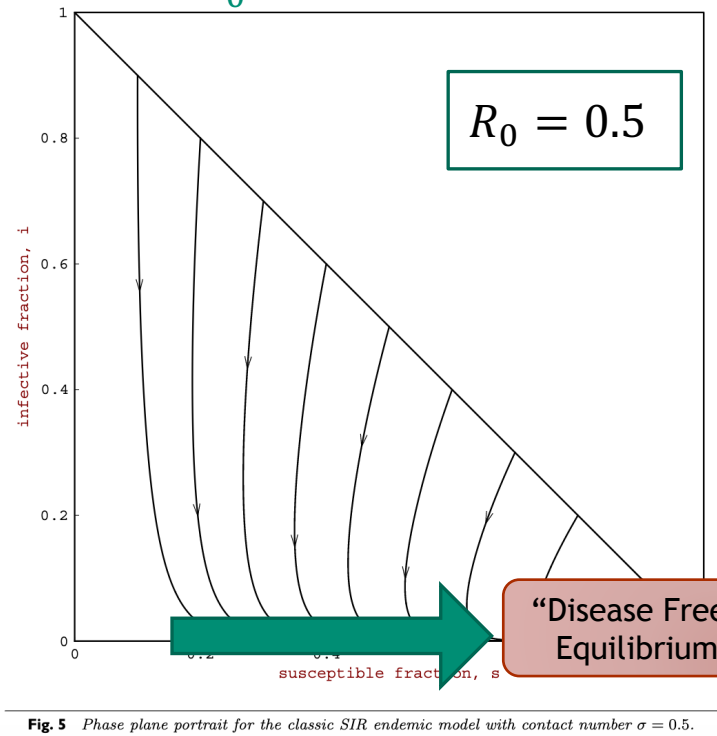


Fig. 5 *Phase plane portrait for the classic SIR endemic model with contact number $\sigma = 0.5$.*

Image Credit: Hethcote 2000

Classic Compartmental Models: Two model forms

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

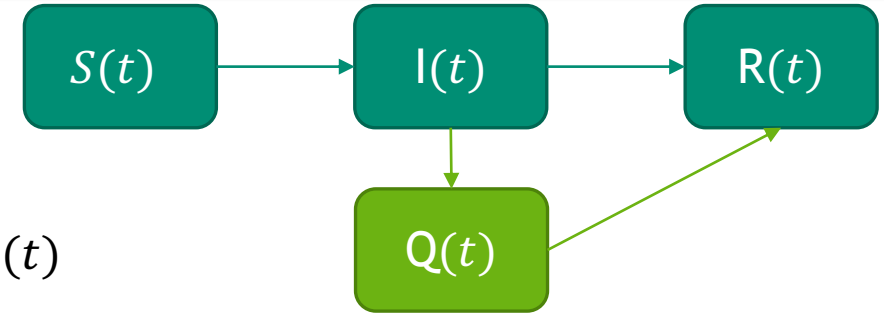
$$\frac{dI}{dt} = \beta \frac{I(t)}{N} S(t) - \gamma I(t)$$

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

$$N = S(t) + I(t) + R(t), \qquad \frac{dN}{dt} = 0$$

$$R_0 = \beta \left(\frac{1}{\gamma}\right)$$

$$R_e(t) = \beta \left(\frac{1}{\gamma}\right) \frac{S(t)}{N}$$



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

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

$$\frac{dQ}{dt} = qI(t) - \gamma_Q Q(t)$$

$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

$$N = S(t) + I(t) + Q(t) + R(t)$$

$$\frac{dN}{dt} = 0$$

$$R_0 = \beta \left(\frac{1}{\gamma_I + q}\right)$$

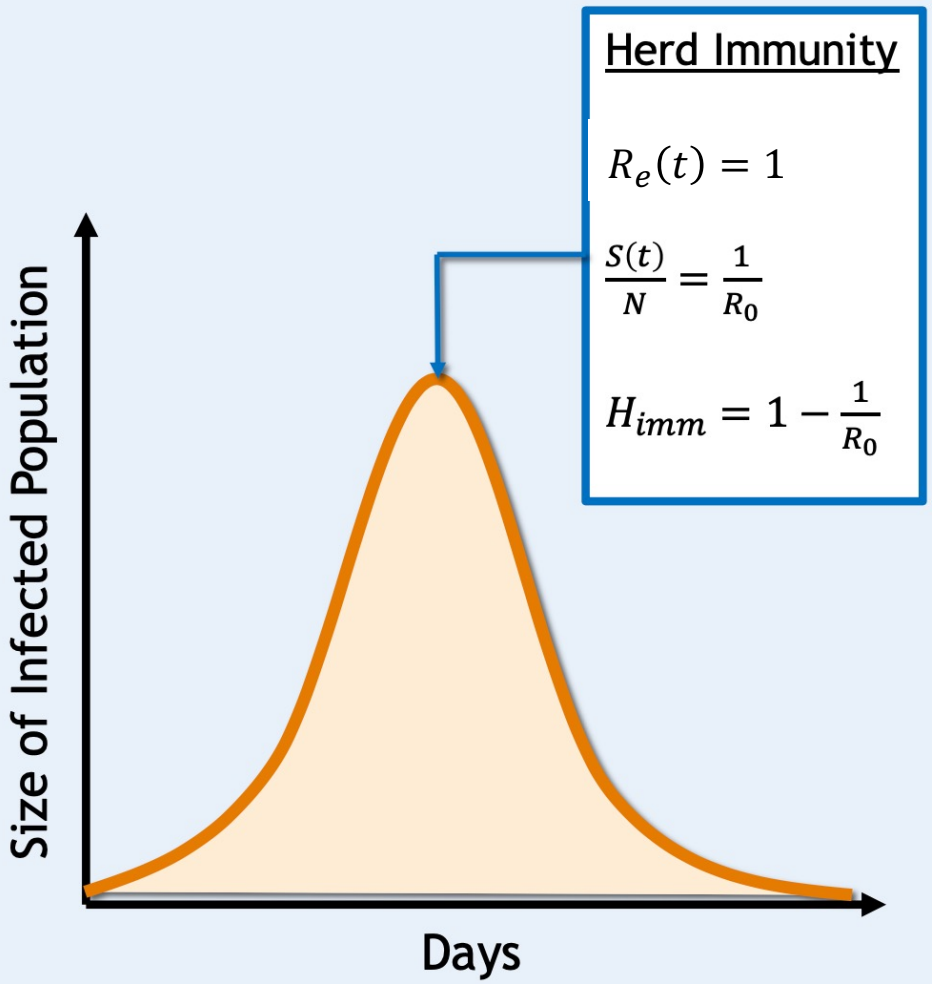
$$R_e(t) = \beta \left(\frac{1}{\gamma_I + q}\right) \frac{S(t)}{N}$$

For these two model forms to generate similar trajectories we require the following conditions:

- $\gamma = \gamma_I + q$: forcing residence time in $I(t)$ to be the same for both models
- $\frac{1}{\gamma_Q} \ll \varepsilon$: requiring that $R(t)$ population results in similar trajectories

Resulting in a model where $Q(t)$ is obsolete

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.

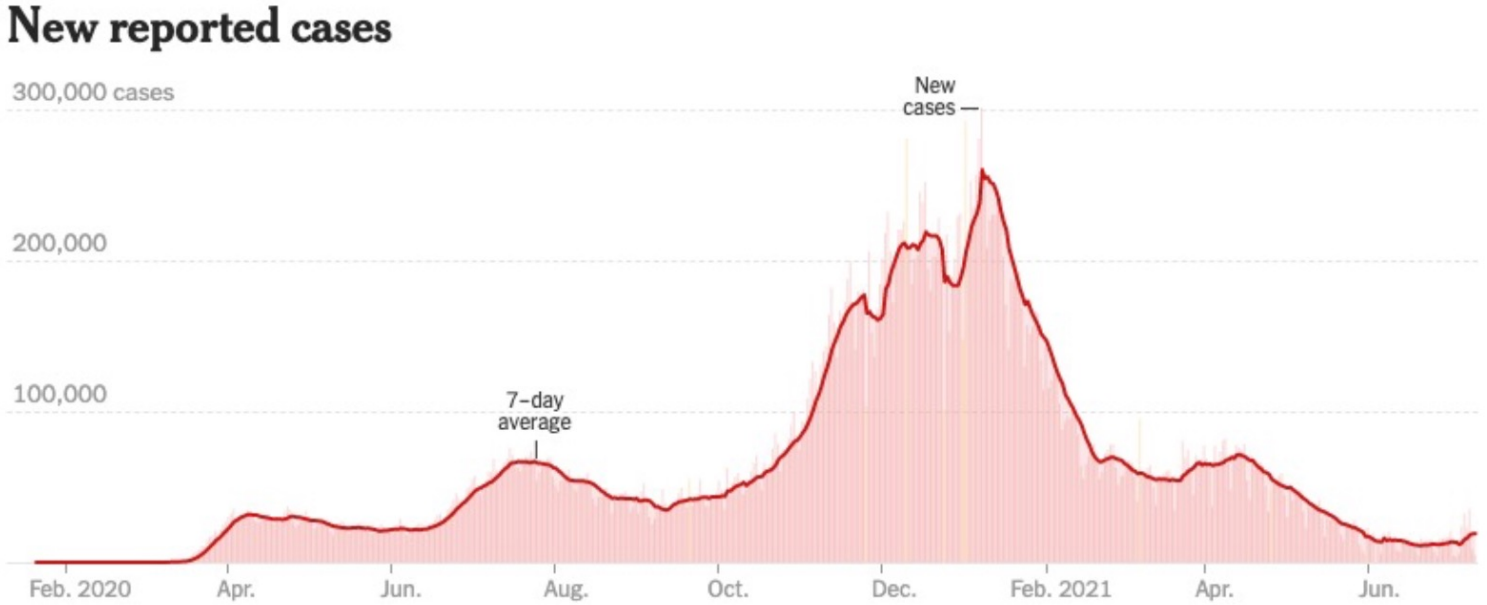
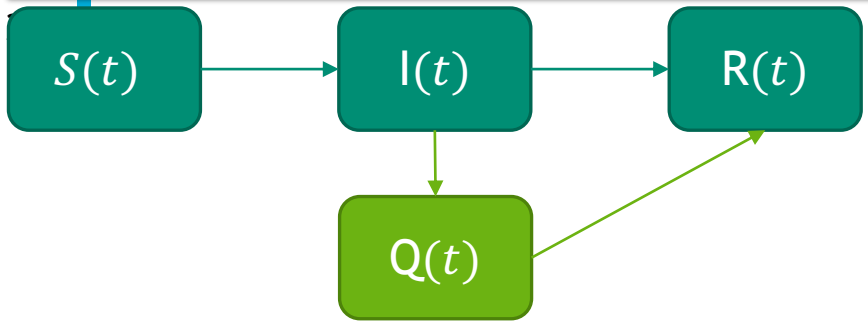


Image Credit: NYT https://www.nytimes.com/interactive/2021/us/covid-cases.html [accessed 2020/07/12]

Regardless of model form, all compartmental model lack the ability to capture the variability of transition rates that are better captured by time dependent transmission rates.

Quarantine Rate as a function of time



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

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

$$\frac{dQ}{dt} = q(t) I(t) - \gamma_Q Q(t)$$

$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

$$N = S(t) + I(t) + Q(t) + R(t)$$

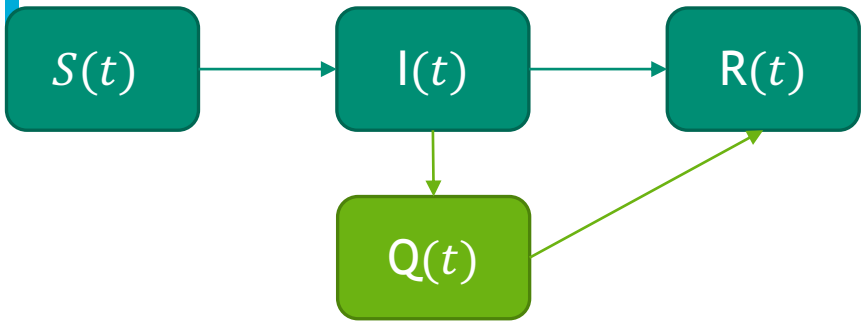
$$\frac{dN}{dt} = 0$$

- If $q(t) = q$ a constant rate for all time, then $R_0 = \beta \left(\frac{1}{\gamma_I + q}\right)$
- A formal analysis is required to verify the bifurcation of the QSIR model with time varying quarantine rate, $q(t)$
- For the purpose of our analysis, we consider the comparison of model solutions with regards to the solutions for the state trajectories as well as the effective reproductive number:

$$R_e(t) = \beta \left(\frac{1}{\gamma_I + q(t)}\right) \frac{S(t)}{N}$$

- We still have $\left(\frac{1}{\gamma_I + q(t)}\right)$ determines the time varying residence time in $I(t)$.

UDEs for compartmental models for infectious disease



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

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

$$\frac{dQ}{dt} = q(t)I(t) - \gamma_Q Q(t)$$

$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

$$N = S(t) + I(t) + Q(t) + R(t)$$

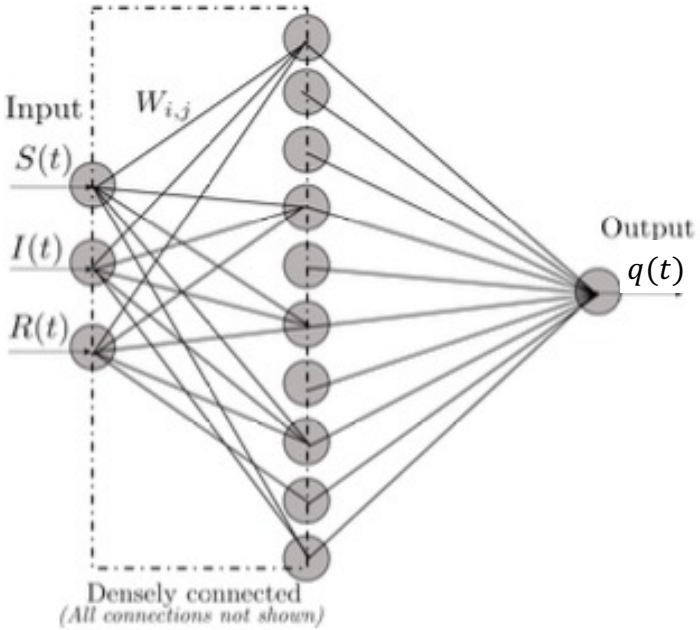
$$\frac{dN}{dt} = 0$$

$$q(t) \approx \mathcal{NN}(SIR, \Theta_{\mathcal{NN}})$$

$$\boldsymbol{u}' = \boldsymbol{F}\left(\boldsymbol{u}, t, \boldsymbol{\Theta}_{ODE}, \mathcal{NN}(\boldsymbol{u}, \boldsymbol{\Theta}_{\mathcal{NN}})\right)$$

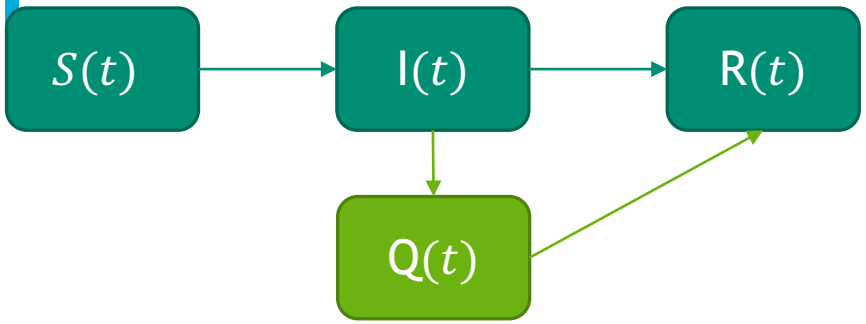
$$\min_{\{\boldsymbol{\Theta}_{ODE}, \boldsymbol{\Theta}_{\mathcal{NN}}\}} \|\boldsymbol{d} - \boldsymbol{u}(\boldsymbol{\Theta}_{ODE}, \boldsymbol{\Theta}_{\mathcal{NN}})\|$$

$\boldsymbol{\Theta}_{\mathcal{NN}} = \{W, b\}$: the weights and biases of the neural network
$\boldsymbol{\Theta}_{ODE} = \{\beta, \gamma_I, \gamma_Q\}$: ODE parameters
$\boldsymbol{d}$: data for observable states



- Neural network architecture:
  - 1 hidden layer, 20 neurons, and a ReLU activation function
  - Densely connected

UDEs for compartmental models for infectious disease



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

$$\frac{dI}{dt} = \beta \frac{I(t)}{N} S(t) - \gamma_I I(t) - \mathcal{NN}(SIR, \Theta_{\mathcal{NN}})\mathrm{I(t)}$$

$$\frac{dQ}{dt} = \mathcal{NN}(SIR, \Theta_{\mathcal{NN}}) I(t) - \gamma_Q Q(t)$$

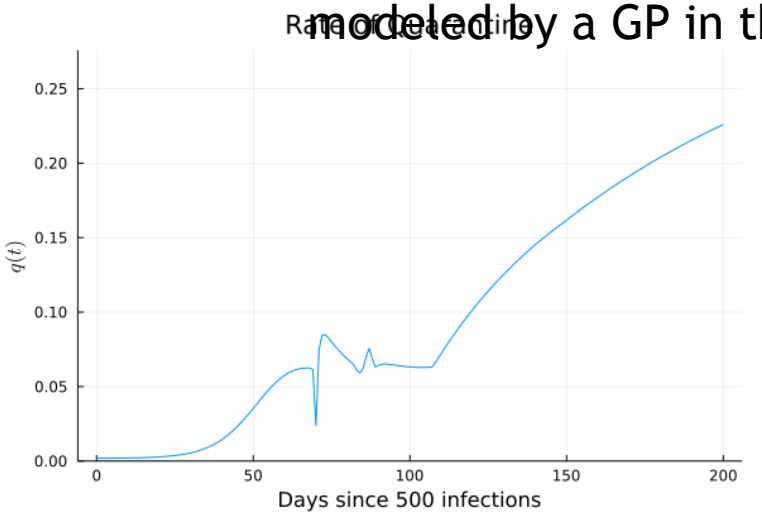
$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

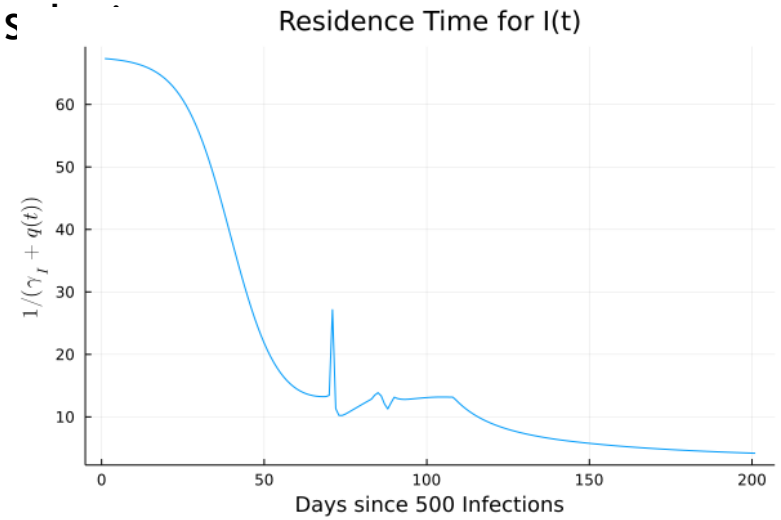
$$N = S(t) + I(t) + Q(t) + R(t)$$

$$\frac{dN}{dt} = 0$$

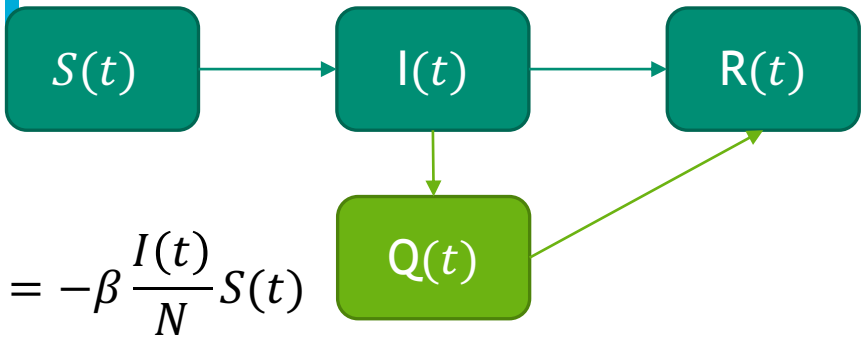
To motivate the discussion on model form and parameter estimation:

- Synthetic data was generated to evaluate K&O approach to Bayesian calibration.
- A neural network was still utilized for $q(t) \approx \mathcal{NN}(SIR, \Theta_{\mathcal{NN}})$
  - 1 hidden layer, 20 neurons, and a Rectified Linear Unit (ReLU) activation function
  - Function plotted below
- $\Theta_{ODE} = \{\beta, \gamma_I, \gamma_Q\} = \{0.15, 0.013, 0.01\}$
- The effects of quarantine has its greatest impact in reducing residence time in $I(t)$.
  - Residence time spreading infection also plotted below.
- K&O was then applied to classic SIR
  - Attempt to learn β and γ_I assuming discrepancy can be modeled by a GP in the s





UDEs for compartmental models for infectious disease



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

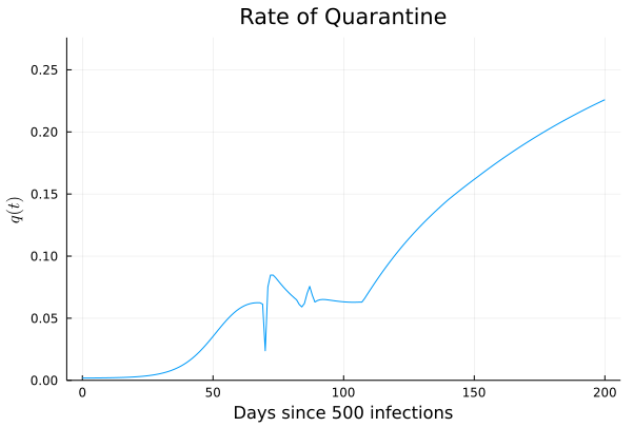
$$\frac{dI}{dt} = \beta \frac{I(t)}{N} S(t) - \gamma_I I(t) - \mathcal{NN}(SIR, \Theta_{\mathcal{NN}})\mathrm{I(t)}$$

$$\frac{dQ}{dt} = \mathcal{NN}(SIR, \Theta_{\mathcal{NN}}) I(t) - \gamma_Q Q(t)$$

$$\frac{dR}{dt} = \gamma_I I(t) + \gamma_Q Q(t)$$

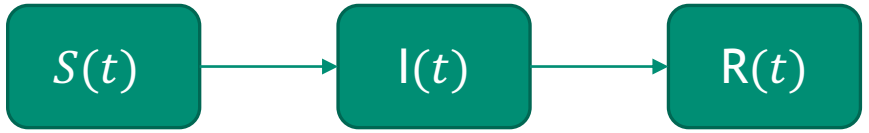
$$N = S(t) + I(t) + Q(t) + R(t)$$

$$\frac{dN}{dt} = 0$$



K&O was then applied to classic SIR

- Attempt to learn β and γ assuming discrepancy can be modeled by a GP in the solution space.



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

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

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

$$N = S(t) + I(t) + R(t),$$

$$\frac{dN}{dt} = 0$$

$$\boldsymbol{y}(t_i) = \boldsymbol{\eta}(t_i, \Theta_{ODE}) + \delta(t_i) + \varepsilon_i$$

- $\boldsymbol{\eta}(t_i, \Theta_{ODE})$: Solution of the state trajectories for $S(t_i), I(t_i)$, and $R(t_i)$
- $\boldsymbol{y}(t_i)$: Synthetic data generate with the QSIR model and a nonlinear transmission rate
  - $[S_{data}(t_i), I_{data}(t_i) + Q_{data}(t_i), R(t_i)]$
- $\delta(t_i)$: Model discrepancy is assumed to follow a Gaussian Process
  - $GP(\mu_\delta, \Sigma_\delta)$: fit to the correlated residuals
  - $\Sigma_\delta = \phi R$
  - $R_{ij} = e^{-\mathrm{K}(t-t_j)^2}$
- ε_i: process error (e.g., noise)
  - $\varepsilon_i = 0$ for all i (i.e., noise-free case)

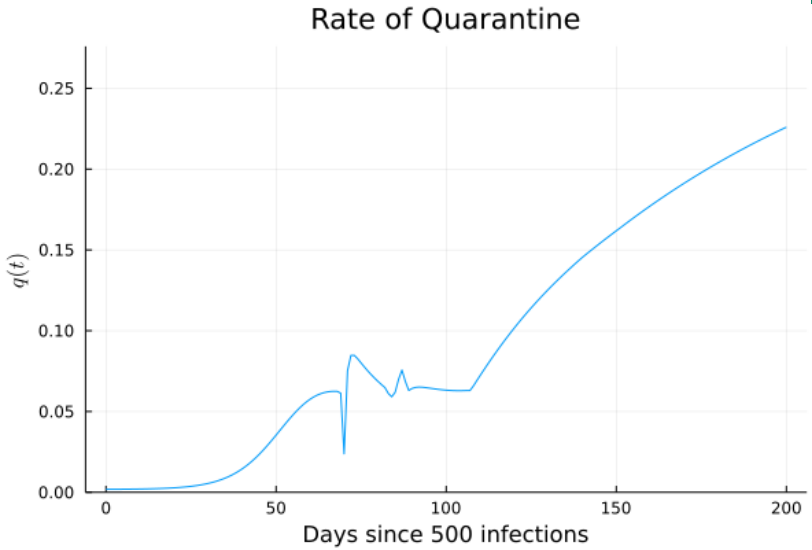
Tracking SIR Model Fit to UDE nonlinear Quarantine Rate $q(t)$

Ground "Truth" Original Parameters for UDE for QSIR:

$\beta = 0.15,\ \gamma = 0.013,\ \delta = 0.1 \quad \Rightarrow R_0 \approx 11.5$

$[S(0), I(0), R(0), Q(0)] \approx [1.999480e6,\ 500,\ 10,\ 10] \quad \Rightarrow \frac{S(0)}{N} \approx 0.99974$

$q(t)$: plotted in "Rate of Quarantine" figure →



Kennedy and O'Hagan parameter estimation for SIR :

$\beta_1 = 0.1350,\ \gamma_1 = 0.0151 \quad \Rightarrow [R_0]_1 \approx 8.9$

$\beta_2 = 0.1476,\ \gamma_2 = 0.0130 \quad \Rightarrow [R_0]_2 \approx 11.3$

$[S(0), I(0), R(0)] \approx [1.999480e6,\ 510,\ 10] \quad \Rightarrow \frac{S(0)}{N} \approx 0.99974$

Close Approximation to $I(t) + Q(t)$ with SIR:

$\beta_3 = 0.1200,\ \gamma_3 = 0.0180 \quad \Rightarrow [R_0]_3 \approx 6.67$

$[S(0), I(0), R(0)]_3 \approx [1.998163e6,\ 1711, 125] \quad \Rightarrow \left[\frac{S(0)}{N}\right]_3 \approx 0.99920072$

Close Approximation to $R_e(t)$ with SIR:

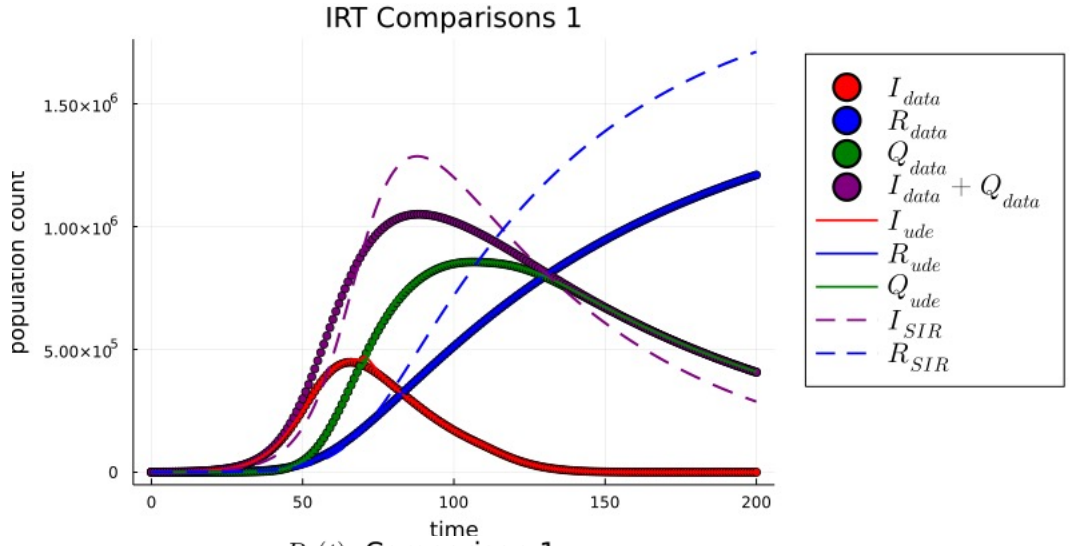
$\beta_4 = 0.1300,\ \gamma_4 = 0.0130 \quad \Rightarrow [R_0]_4 = 10$

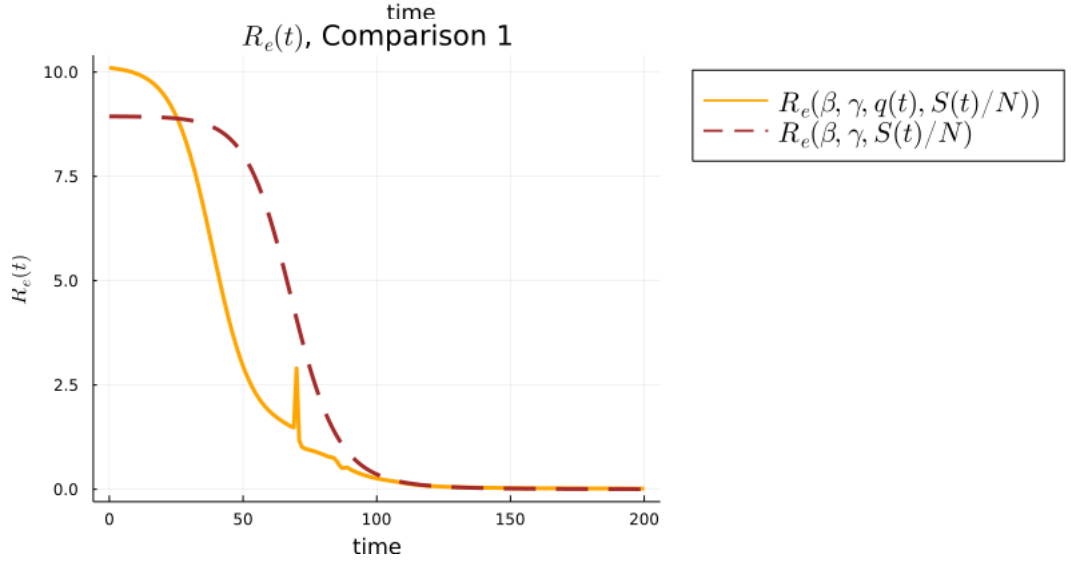
$[S(0), I(0), R(0)]_4 \approx [1.992834e6,\ 6572, 593] \quad \Rightarrow \left[\frac{S(0)}{N}\right]_4 \approx 0.99768227$

Kennedy and O'Hagan parameters estimations for SIR :

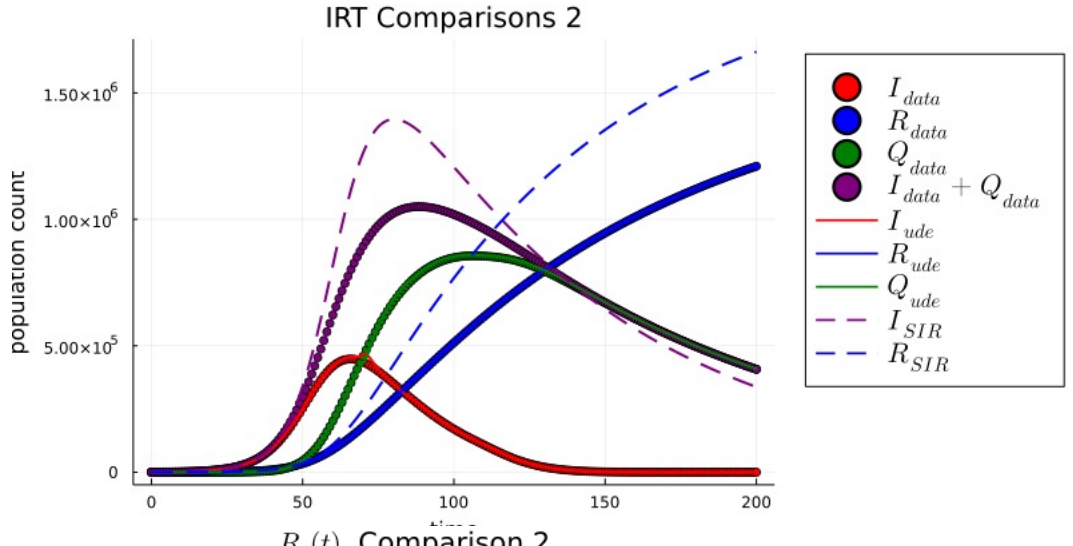
$[S(0), I(0), R(0)] \approx [1.999480e6,\ 510,\ 10] \quad \Rightarrow \frac{S(0)}{N} \approx 0.99974$

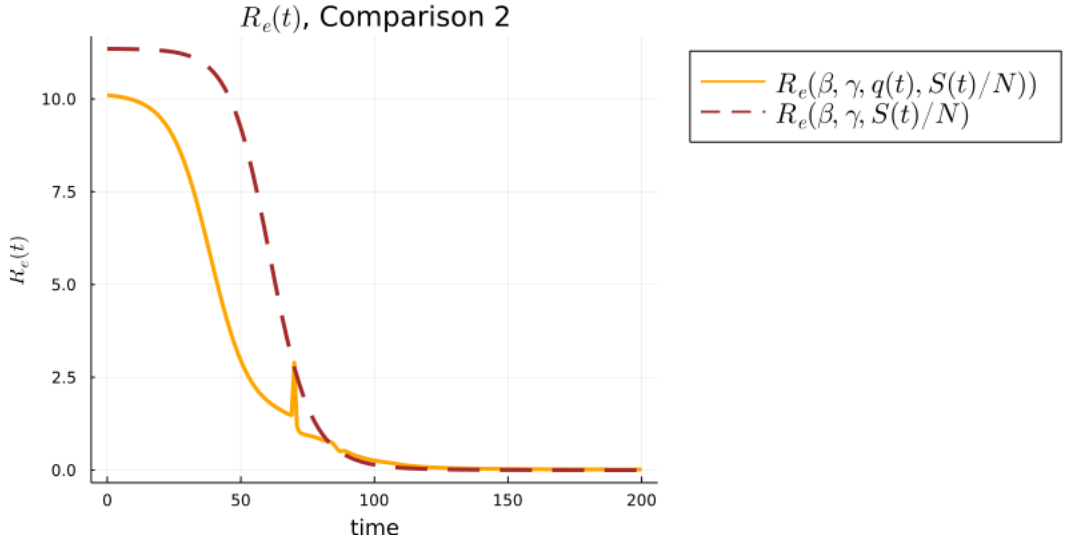
$\beta_1 = 0.1350,\ \gamma_1 = 0.0151 \quad \Rightarrow [R_0]_1 \approx 8.9$





$\beta_2 = 0.1476,\ \gamma_2 = 0.0130 \quad \Rightarrow [R_0]_2 \approx 11.3$



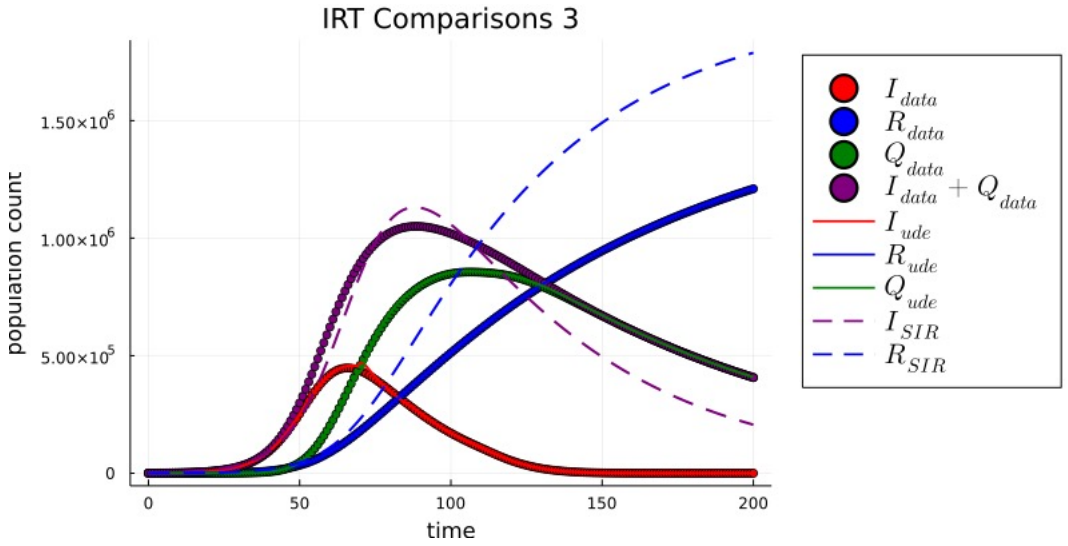


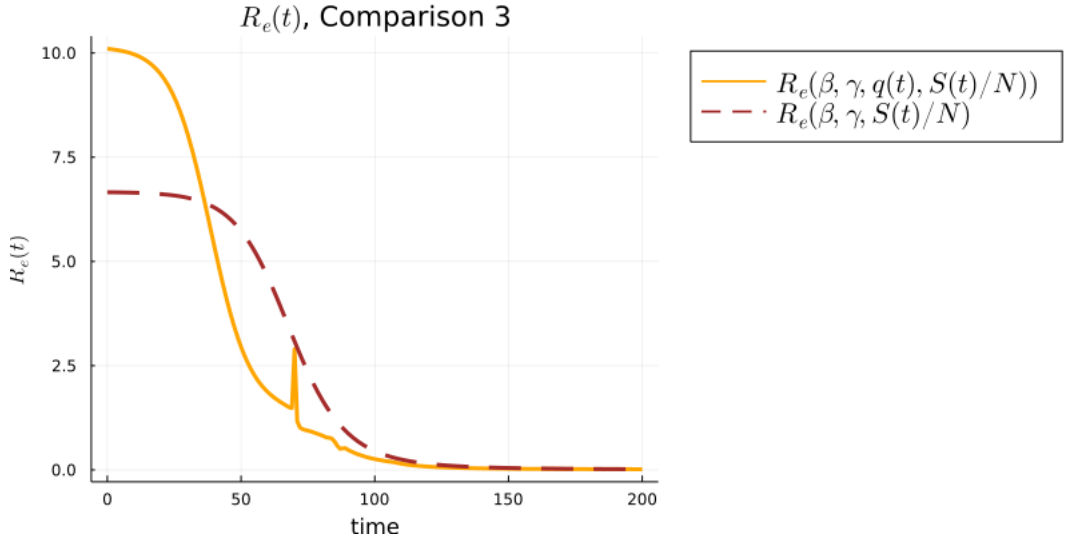
Close Approximations Allowing Initial State to change:

$\beta_3 = 0.1200,\ \gamma_3 = 0.0180 \quad \Rightarrow [R_0]_3 \approx 6.67$

$[S(0), I(0), R(0)]_3 \approx [1.998163e6,\ 1711, 125]$

$\Rightarrow \left[\frac{S(0)}{N}\right]_3 \approx 0.99920072$

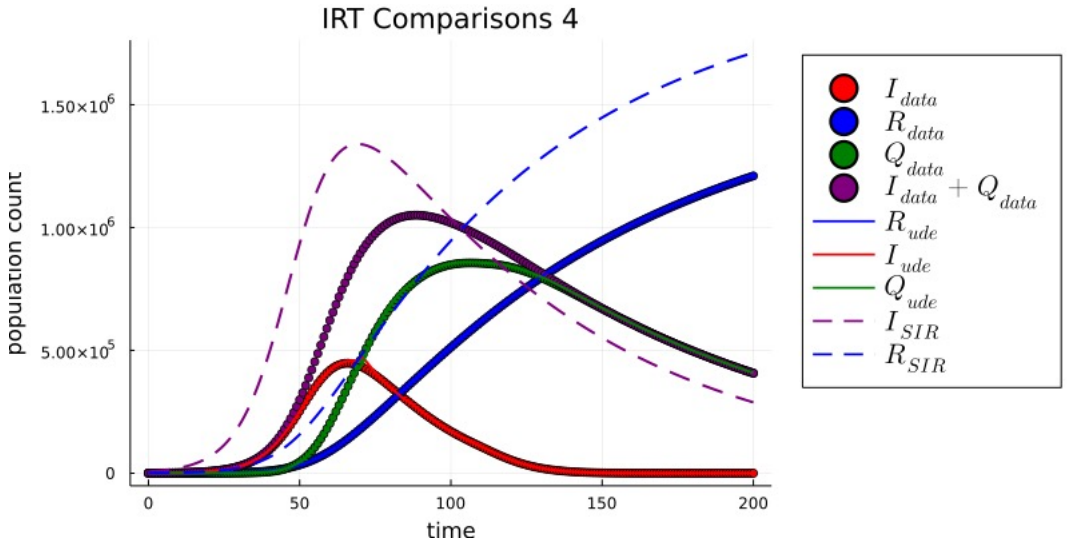


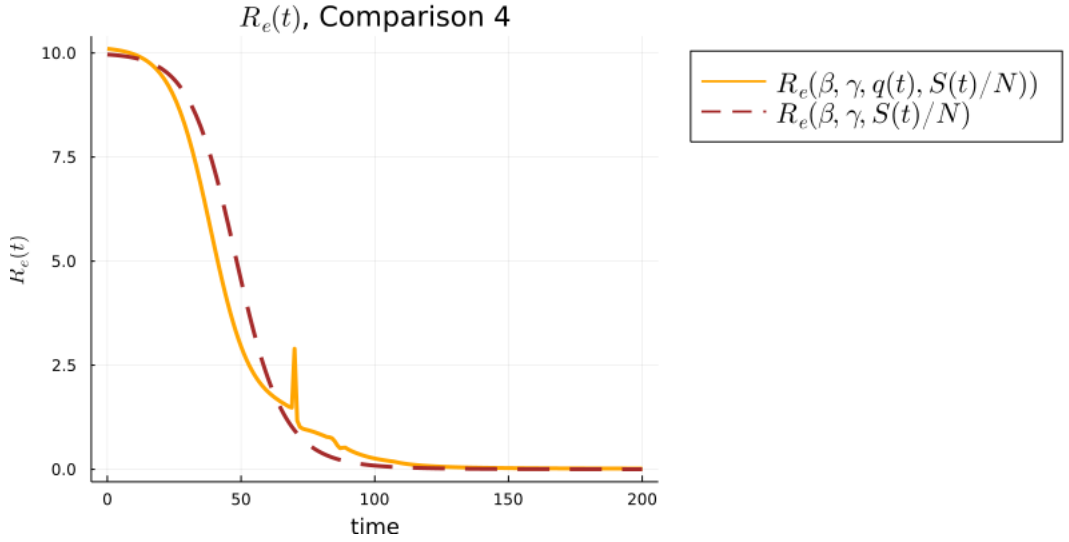


$\beta_4 = 0.1300,\ \gamma_4 = 0.0130 \quad \Rightarrow [R_0]_4 = 10$

$[S(0), I(0), R(0)]_4 \approx [1.992834e6,\ 6572, 593]$

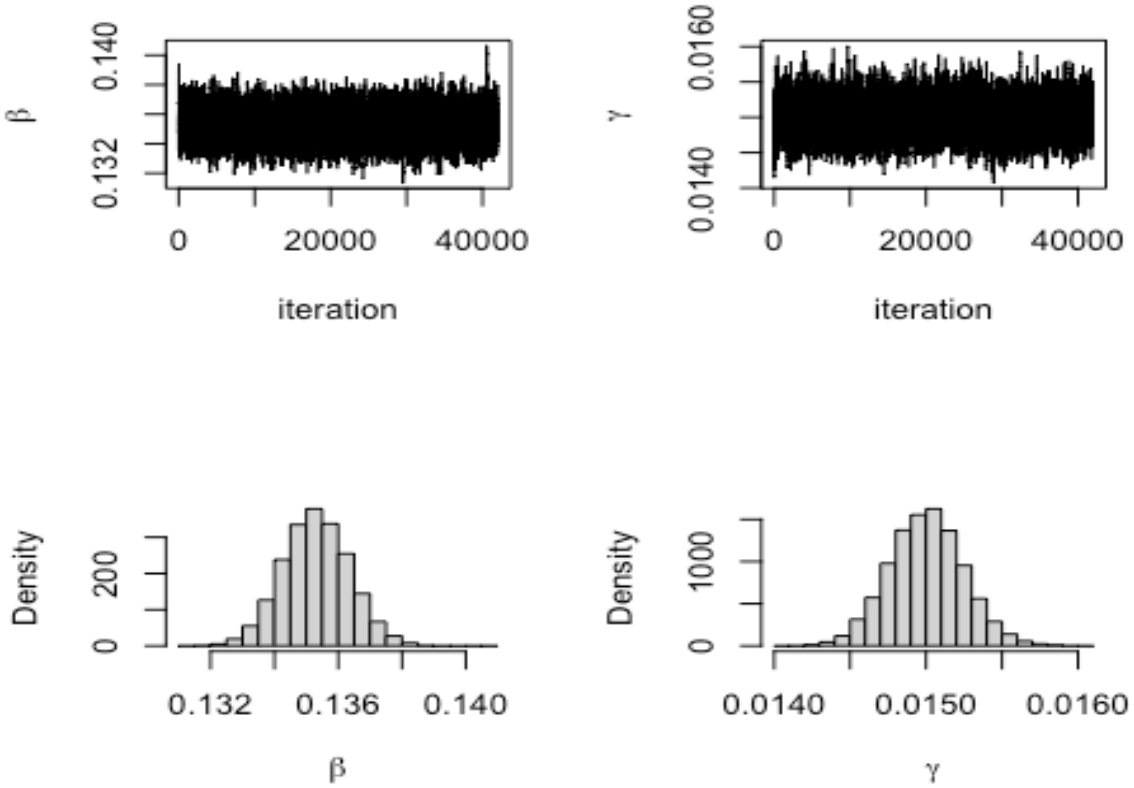
$\Rightarrow \left[\frac{S(0)}{N}\right]_4 \approx 0.99768227$



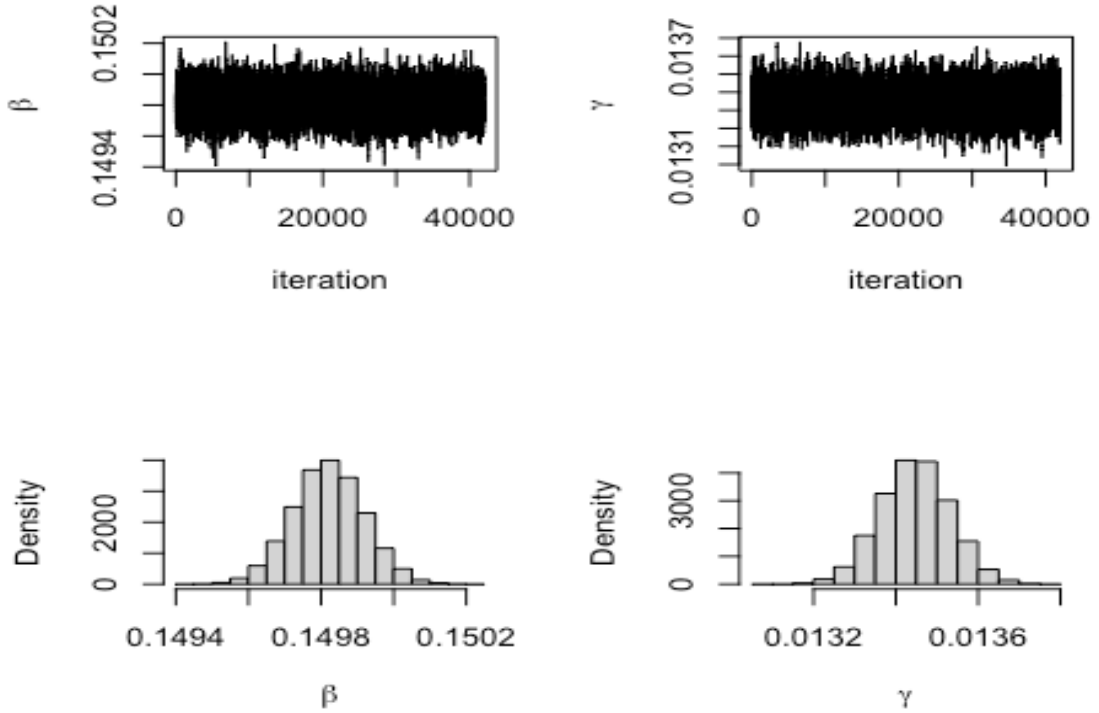


Results of K&O Bayesian Calibration

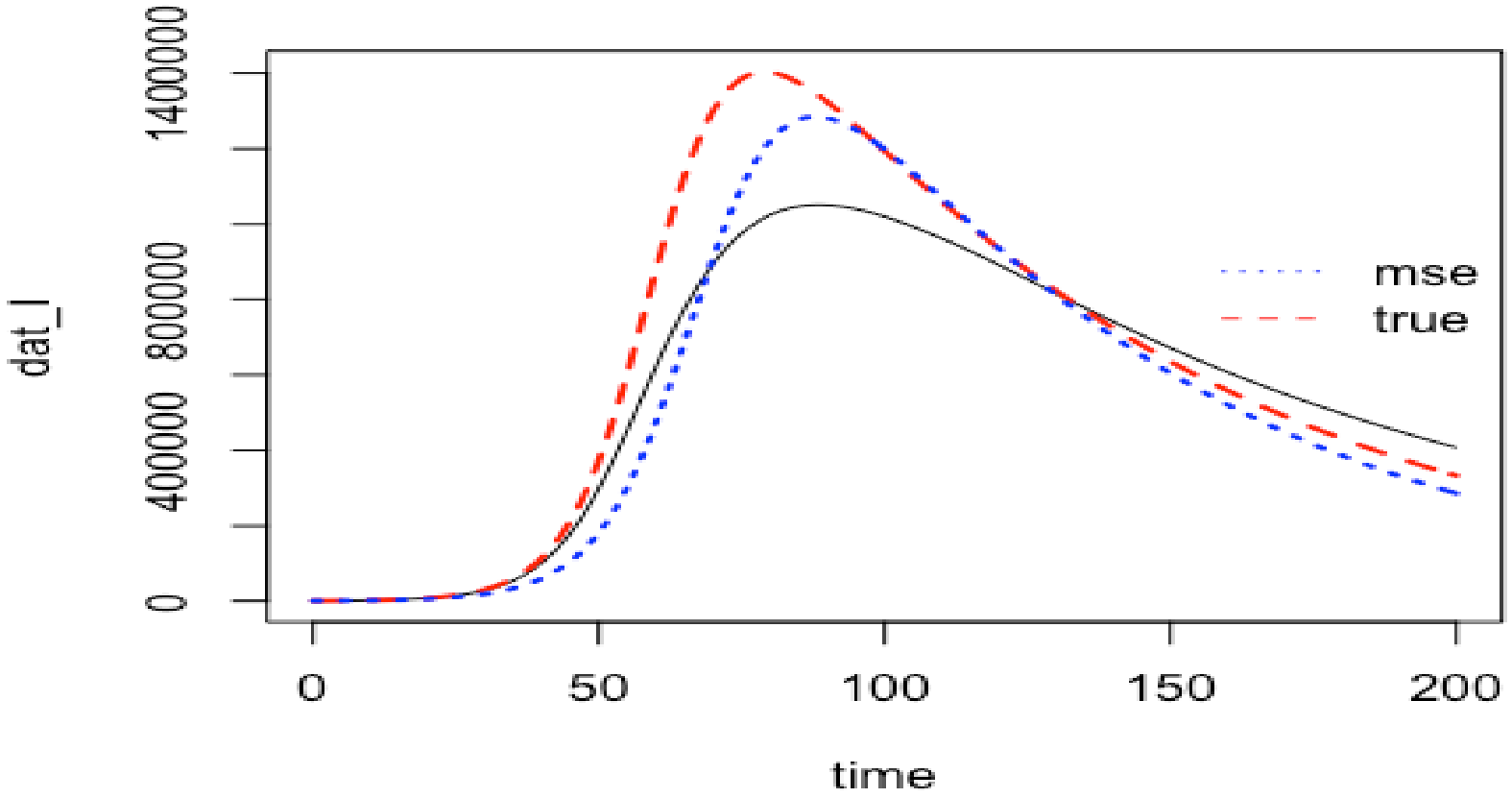
Result with “weak” prior centered at



Result with ”strong” prior

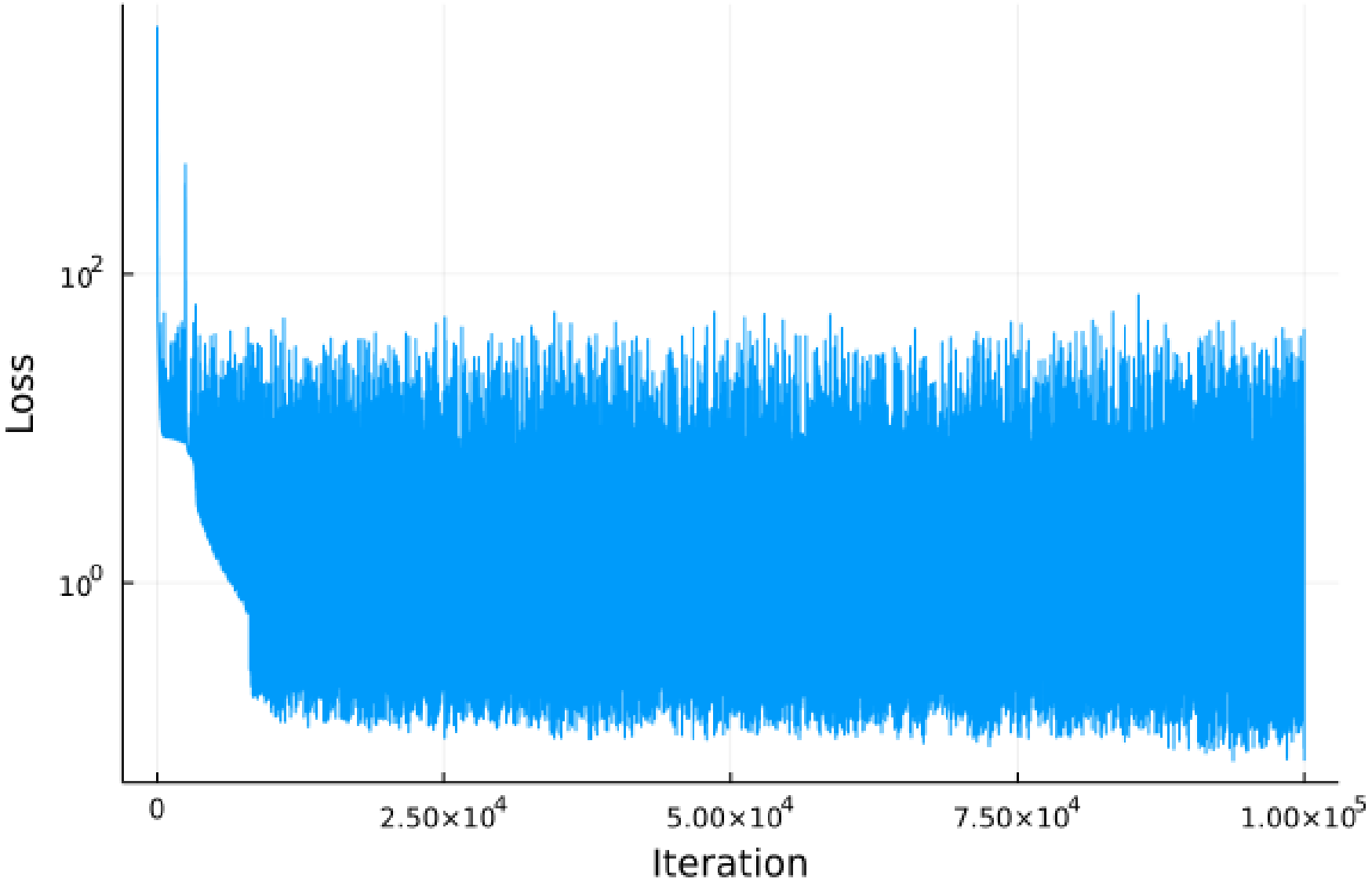


Comparison of resulting solutions w/o discrepancy



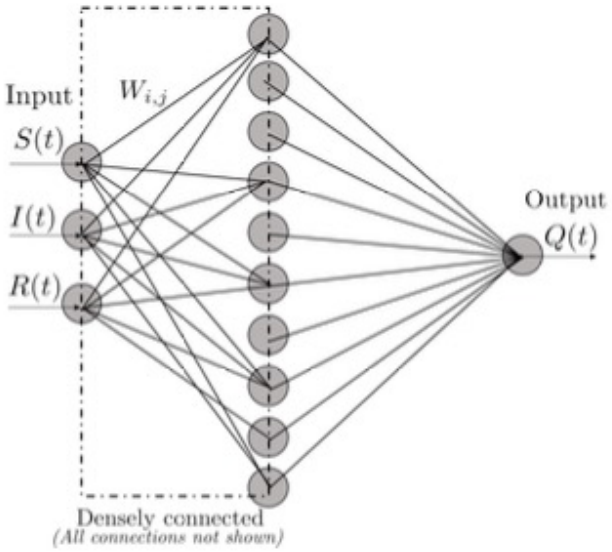
- Black line corresponds to the data
- Dashed red is the solution from SIR with the “true” values: 0.15,0.013.
- Dotted blue is a solution based on optimization of squared-error-loss.

Example of neural network loss function realizations for each iteration of the optimization algorithm



UDEs

Key Theoretical Foundation of SciML



Universal Approximation Theorem [[11],[12],[13]]:

(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

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

While the Universal Approximation Theorem is a necessary condition for neural networks to be function approximators, in practice this is not a sufficient condition.

Ensemble Training: Robust Learning and Uncertainty Quantification

Approach:

For each combination: $\{[I, R, Q], [I, R], [I, Q], [R, Q], [I], [R], [Q]\}$

Initialize model parameters $\Theta = \{\Theta_{ODE}, \Theta_{NN}\}$

a. Θ_{ODE} sampled from distributions derived from the literature.
b. Θ_{NN} established from Glorot initialization

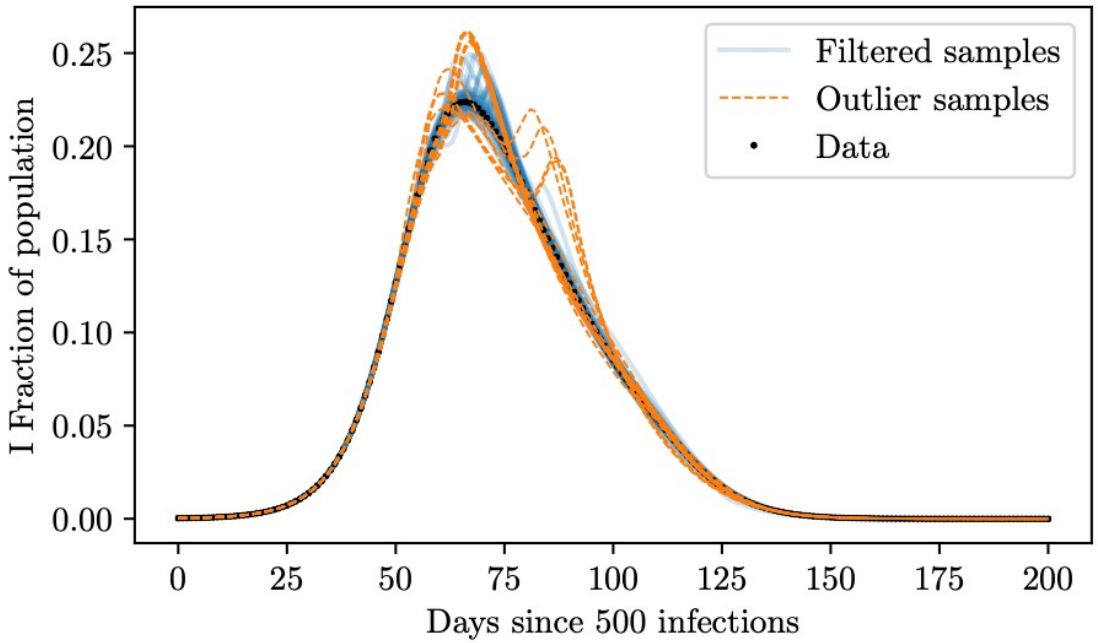
Run 100 training replicates to learn: $\{\widehat{\Theta}_{NN}^{k}\}$ and $\{\widehat{\Theta}_{ODE}^{k}\}$, for $k = 1, \ldots, 100$.

Challenge:

Optimization is sensitive to initialization and can get stuck in a local minima.

Mitigation:

Filter out outlier ensemble members (those with very large MSE).



Noisy Data Generation

Stochastic Differential Equations for Noisy Data Generation

$$\frac{dS}{dt} = -\lambda(t)S(t)$$

$$\frac{dI}{dt} = \lambda(t)S(t) - \mu_{\gamma I}I(t) - \sigma_{\gamma I}^2 I(t) - q(t)I(t)$$

$$\frac{dR}{dt} = \mu_{\gamma I}I(t) + \sigma_{\gamma I}^2 I(t) + \mu_{\gamma Q}Q(t) + \sigma_{\gamma Q}^2 Q(t)$$

$$\frac{dQ}{dt} = q(t)I(t) - \mu_{\gamma Q}Q(t) - \sigma_{\gamma Q}^2 Q(t)$$

$$\beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$$

$$\gamma_I \sim \mathcal{N}(\mu_{\gamma I}, \sigma_{\gamma I}^2)$$

Such that:

$$\lambda(t) = \mu_\beta \frac{I(t)}{S(t) + I(t) + R(t) + Q(t)} + \sigma_\beta^2 \frac{I(t)}{S(t) + I(t) + R(t) + Q(t)}$$

$$q(t) = NN(IRQ; W, b)$$