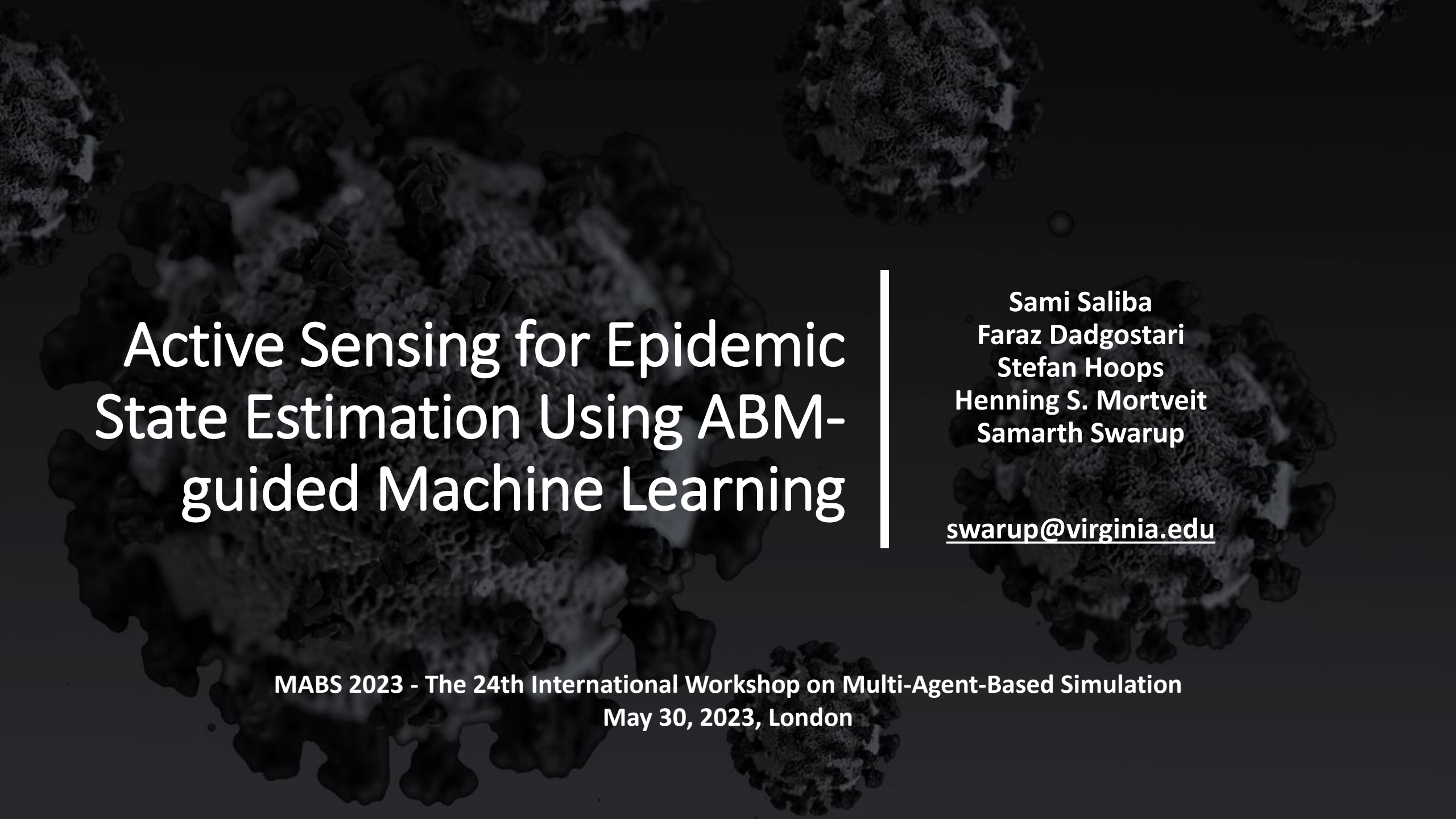


Active Sensing for Epidemic State Estimation Using ABM-guided Machine Learning

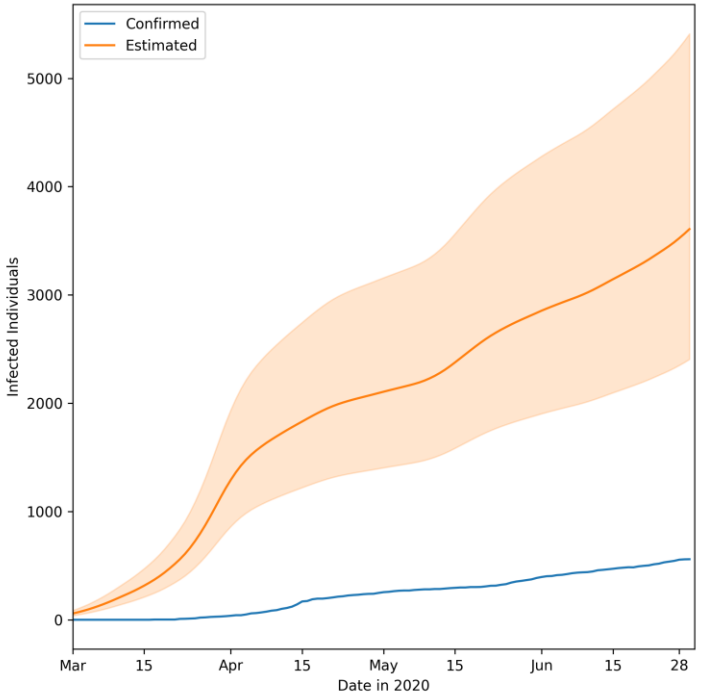
Sami Saliba
Faraz Dadgostari
Stefan Hoops
Henning S. Mortveit
Samarth Swarup

swarup@virginia.edu

MABS 2023 - The 24th International Workshop on Multi-Agent-Based Simulation
May 30, 2023, London

Epidemic State Estimation

- During an epidemic, the true number of infectious people at any time is unknown.
  - Lack of adequate testing
  - Mild symptoms go unreported
  - Misdiagnoses
- Can we develop a method to estimate the true number of infections in an ongoing epidemic?

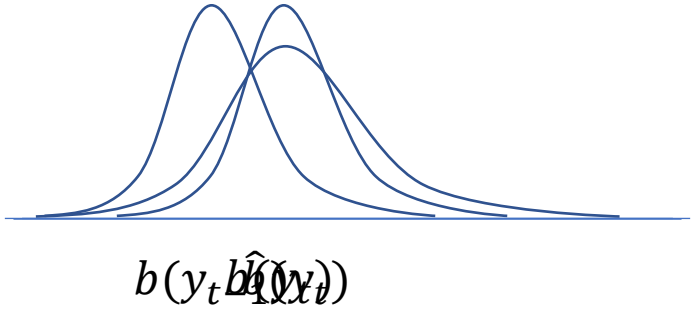


https://covid19-projections.com/

Standard Approach: Bayes Filtering

- Let y_t be the unobserved true state of a dynamical system at time t.
- Since we can't observe it, we would like to maintain a *belief state* that integrates our prior estimate, y_{t-1}, and an observation, o_t.
- A belief state is a posterior distribution over y_t. It is generated in two steps:
  - Prediction: $\hat{b}(y_t) = \int p(y_t|y_{t-1})b(y_{t-1})dy_{t-1}$.
  - Correction: $b(y_t) = \mu p(o_t|y_t)\,\hat{b}(y_t)$.

Note that, in this formulation, we don't control the observation process.



Testing During Epidemics

- Common approaches include targeted testing and contact tracing.
- Prevalence testing is done in controlled settings
  - E.g., military bases, universities
- Prevalence testing of the general population is expensive.
  - If we do this, we would like to do as little of it as possible.

Biased in various unknown ways

From Filtering to Active Sensing

- Our idea is to use a filtering approach, but to control the observation process.
- We would like to use a small number of tests, allocated adaptively, to generate a good estimate of the true number of infections.

$\boldsymbol{y}_t = [y_{0,t}, y_{1,t}, \ldots, y_{N,t}]$: True proportions of infectious people at time t.
$\boldsymbol{x}_t = [x_{0,t}, x_{1,t}, \ldots, x_{N,t}]$: Our estimate of $\boldsymbol{y}_t$.
$\boldsymbol{n}_t = [n_{0,t}, n_{1,t}, \ldots, n_{N,t}], 1 \leq n_{i,t} \leq M$: Numbers of tests assigned at time t.

Intuition: Minimize $R(t) = |\boldsymbol{y}_t - \boldsymbol{x}_t| + \boldsymbol{n}_t$

Problem Formulation

$\boldsymbol{y}_t = [y_{0,t}, y_{1,t}, \ldots, y_{N,t}]$: True proportions of infectious people at time t.
$\boldsymbol{x}_t = [x_{0,t}, x_{1,t}, \ldots, x_{N,t}]$: Our estimate of $\boldsymbol{y}_t$.
$\boldsymbol{n}_t = [n_{0,t}, n_{1,t}, \ldots, n_{N,t}], 1 \leq n_{i,t} \leq M$: Numbers of tests assigned at time t.

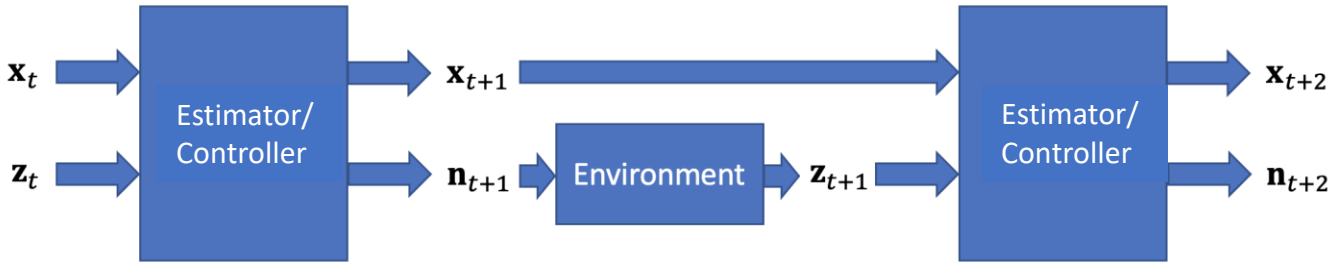
Intuition: Minimize $R(t) = |\boldsymbol{y}_t - \boldsymbol{x}_t| + \boldsymbol{n}_t$

However, $\boldsymbol{y}_t$ is never observed. We only observe test results.

$$Environment: \boldsymbol{n}_t \rightarrow \boldsymbol{z}_t$$

We change our objective function to,

$$r(t) = |\boldsymbol{z}_t - \boldsymbol{x}_t| + \eta|\boldsymbol{n}_t|$$



The estimator/controller generates both the estimate of the true number of infections and the numbers of tests to assign to each region in the next timestep.

Learning Model

- A particle filter maintains the belief state as a collection of particles, $\{y^i\}_t$, with associated weights, $\{w^i\}_t$. In a PF-RNN these are maintained internally.
- During training, the PF-RNN learns the transition model and the observation model.
- We can also augment the objective function with the entropy of the belief state to minimize uncertainty in the estimate.

$$r(t) = |\boldsymbol{z}_t - \boldsymbol{x}_t| + \eta|\boldsymbol{n}_t| + H(\boldsymbol{h}_t)$$

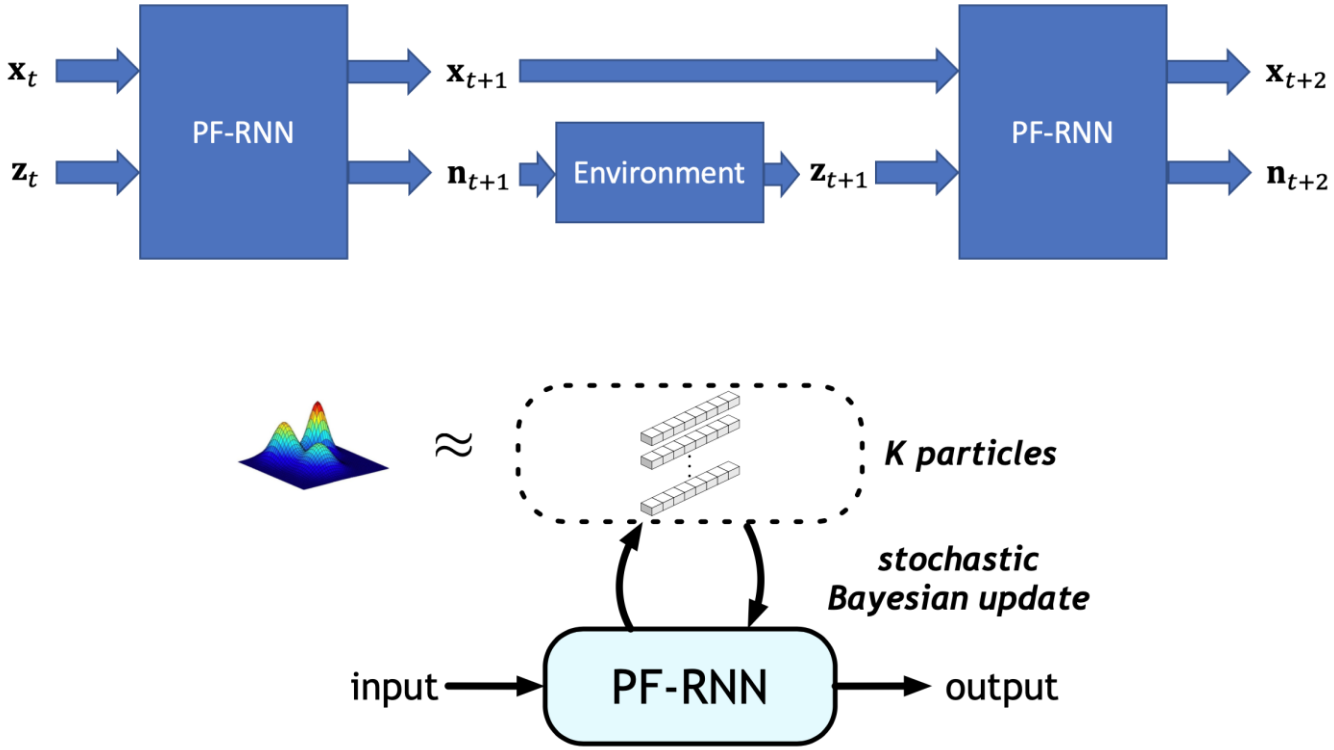


Fig. 1 from Ma et al., AAAI 2020.

Particle Filter Recurrent Neural Networks,
Xiao Ma, Peter Karkus, David Hsu, Wee Sun Lee,
Proceedings of the AAAI Conference on Artificial Intelligence, *34(04)*, 5101-5108, 2020.

Approach

- Use a detailed data-driven ABM to train the model.
- The ABM uses a *synthetic population* of Virginia for realism.

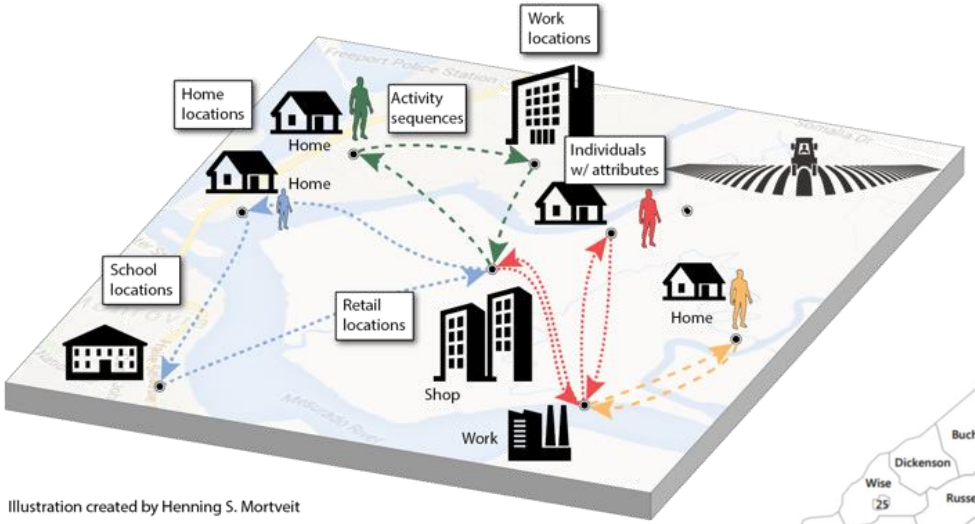
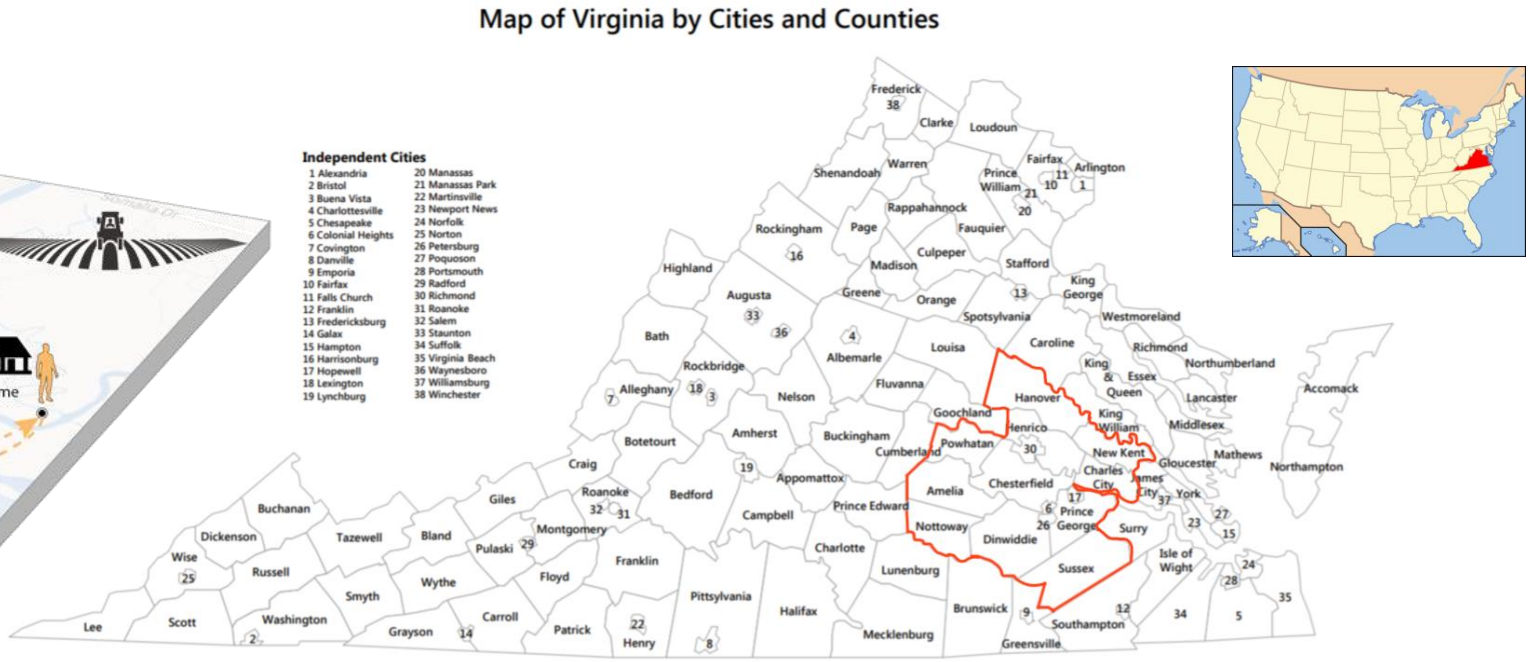


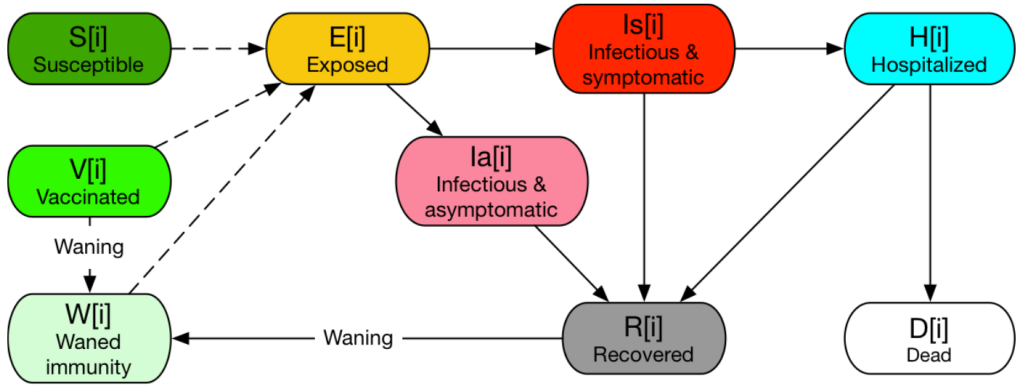
Illustration created by Henning S. Mortveit



Generating a Synthetic Population of the United States
Adiga et al., Network Dynamics and Simulation Science Laboratory, Tech Report NDSSL 15-009, Jan 2015.
https://nssac.bii.virginia.edu/~swarup/papers/US-pop-generation.pdf

The Agent-based Model: EpiHiper

- Works on a labeled, time-varying social contact network derived from the synthetic population.
- Takes duration of contact into account when computing infections.
- Highly parallel, runs on a large cluster.
- Can implement several kinds of interventions.



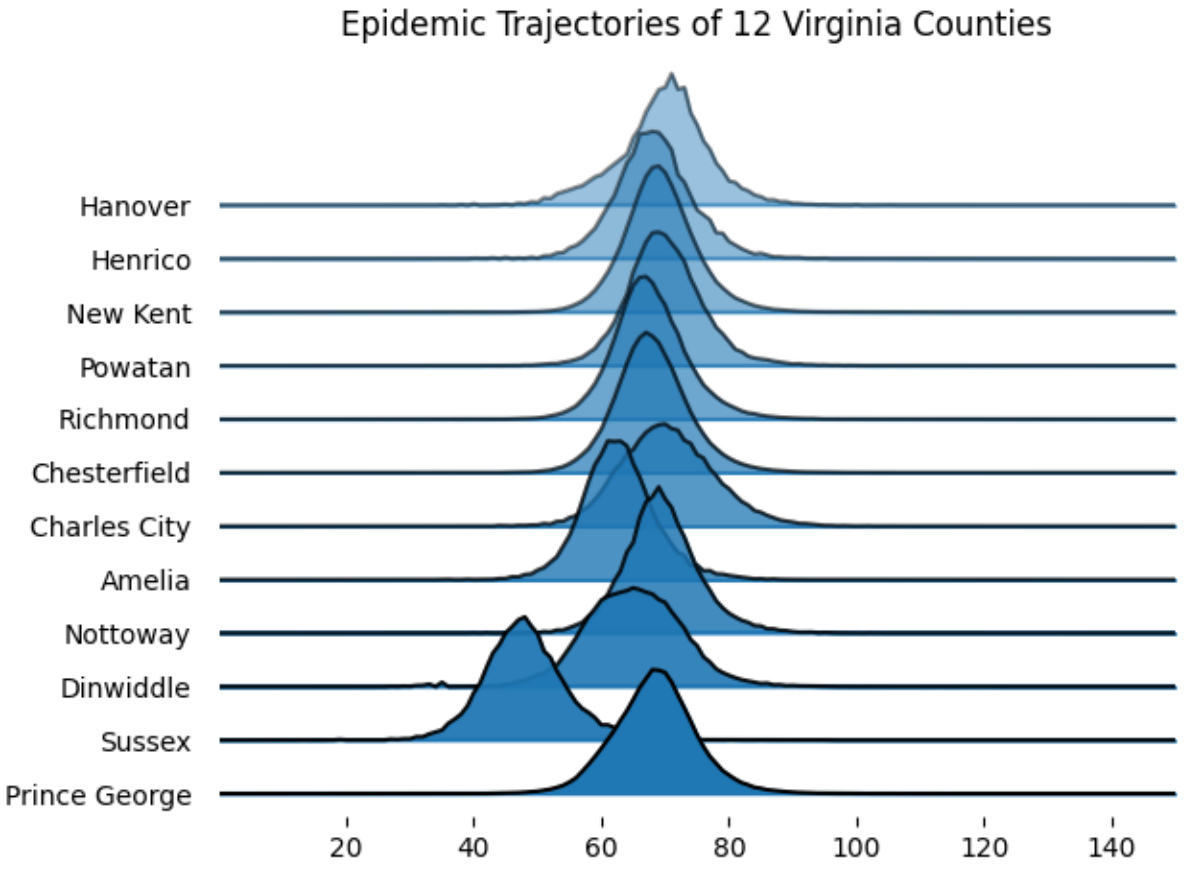
Disease model, age-stratified:

- Preschool: 0 – 4 years
- Students: 5 – 17 years
- Adults: 18 – 49 years
- Older Adults: 50 – 64 years
- Seniors: 65+ years

Scalable epidemiological workflows to support COVID-19 planning and response.
D. Machi, P. Bhattacharya, S. Hoops, J. Chen, H. Mortveit, S. Venkatramanan, B. Lewis, M. Wilson, A. Fadikar, T. Maiden, C.L. Barrett, M. V. Marathe,
Proceedings of the IEEE International Parallel and Distributed Processing Symposium (IPDPS) 2021 (pp. 639-650).

Training Data

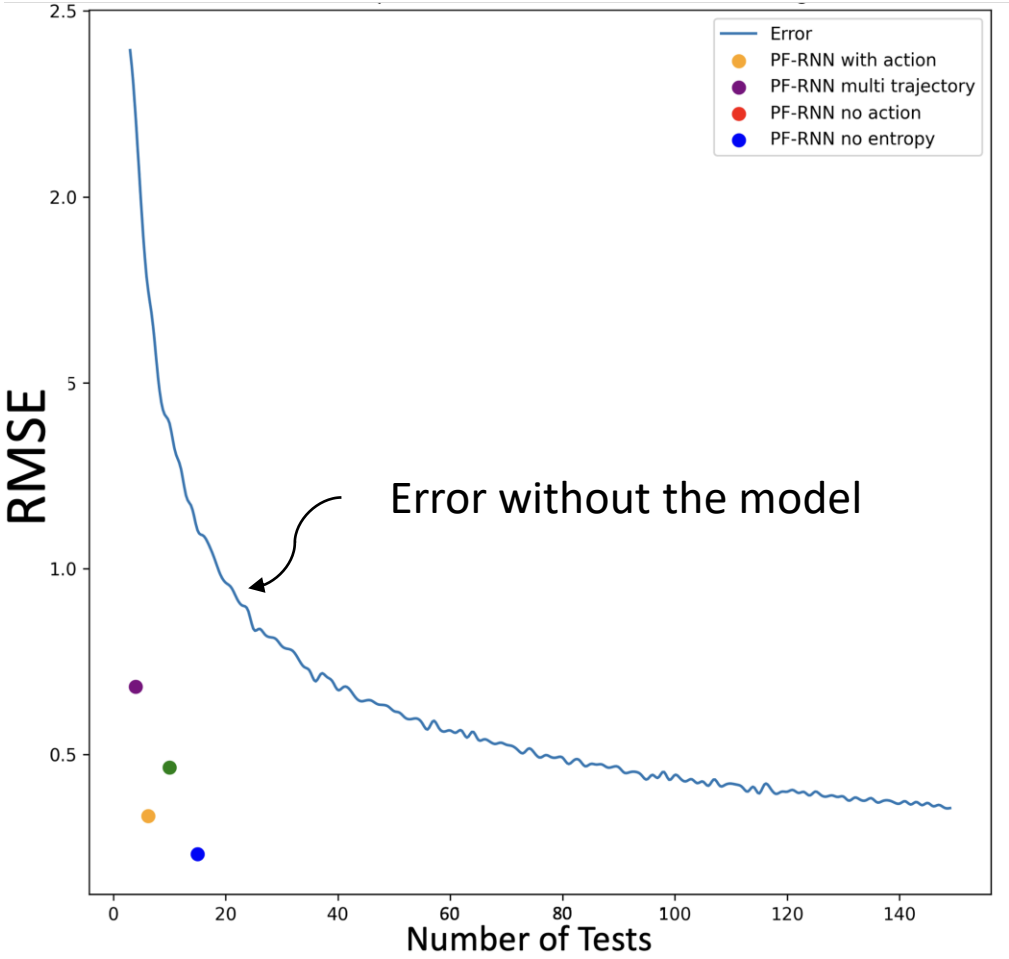
- We generated 10,000 epidemic trajectories for all of Virginia and selected 12 contiguous counties for training the model.
- Multiple unique transmissibility values in the range [0.025,0.095], 50 trajectories per transmissibility value.
- Each trajectory is of length 180 days.

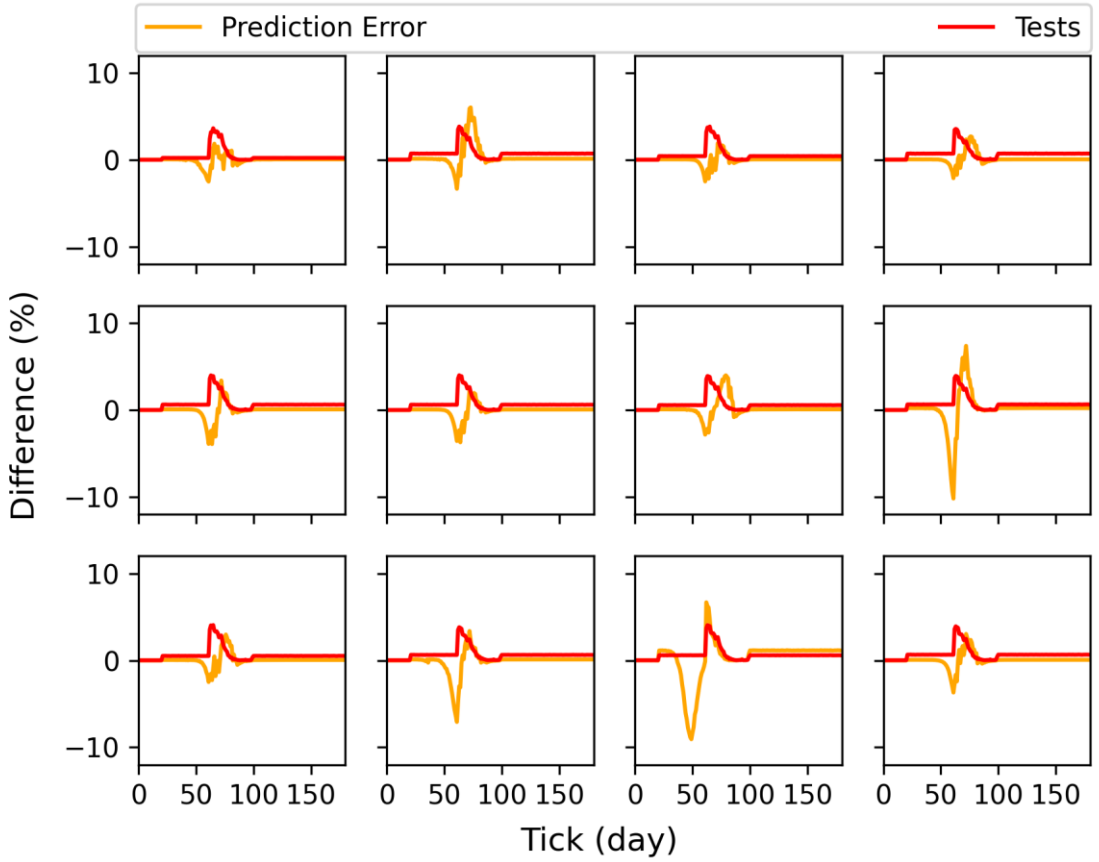


Experiments

- **No action**: 5 tests assigned to each county.
- **With action**: The neural network has to generate both the estimate of the epidemic state at time t and the number of cases to assign to each region in time step t + 1.
- **No entropy**: This is the same as the previous case, but without the term for the variance of the set of particles in the objective function.
- **Multi-trajectory**: In this case, training was done on a multiple trajectories with transmissibility rates in the range of [0.025 – 0.095] and evaluation was conducted on a trajectory with an unseen transmissibility value within this range.

Results


Error
PF-RNN with action
PF-RNN multi trajectory
PF-RNN no action
PF-RNN no entropy
RMSE
Error without the model
Number of Tests
2.5
2.0
5
1.0
0.5
0
20
40
60
80
100
120
140


Prediction Error
Tests
Difference (%)
10
0
−10
0
50
100
150
Tick (day)

Thank you!

Samarth Swarup
University of Virginia
swarup@virginia.edu