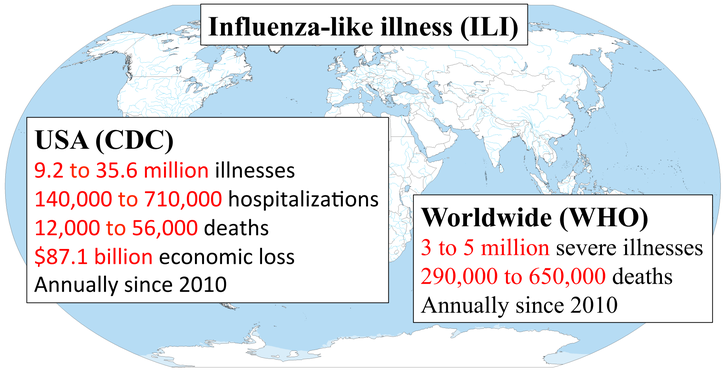
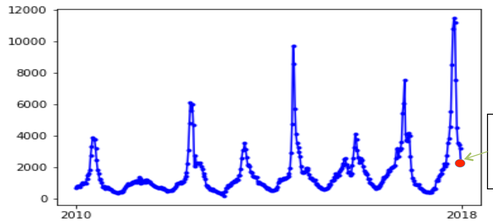
TDEFSI: Theory Guided Deep Learning Based Epidemic Forecasting with Synthetic Information

Lijing Wang[1,2], Jiangzhuo Chen[2], Madhav Marathe[1,2]
1 Department of Computer Science, UVA
2 Network Systems Science and Advanced Computing, Biocomplexity Institute and Initiative, UVA

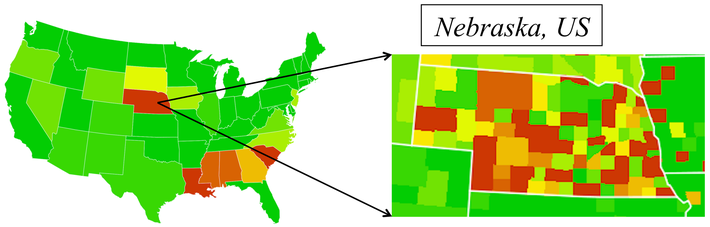
Problem





Data driven methods

Make forecasts at time point t



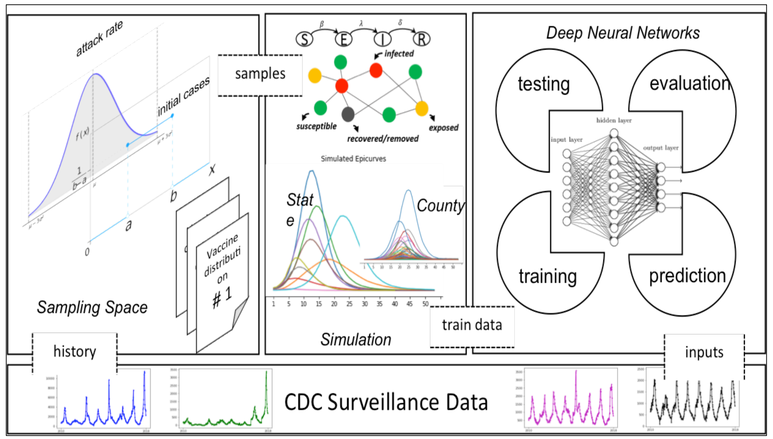
Flat-resolution (state level) surveillance data

High-resolution (county level) forecasts

Proposed Method:

TDEFSI = Causal Model + DNN Model

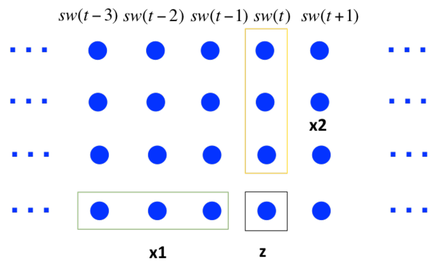
Framework and Models



I. Disease model parameter space learning | **II. Synthetic training data generating** | **III. DEFSI training and predicting**

1. Agent-based SEIR disease model
2. LSTM-based models incorporating physical consistency

$$\hat{\mathbf{z}}_t = f([\mathbf{x1}, \mathbf{x2}]_t, \theta)$$



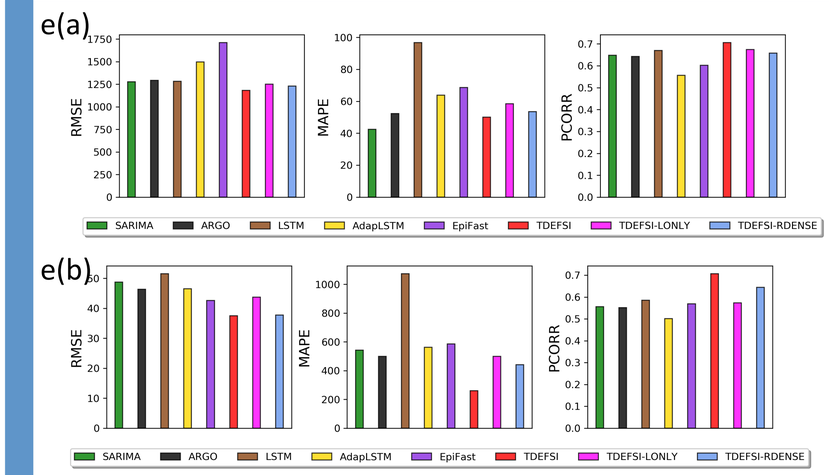
Optimization objective:

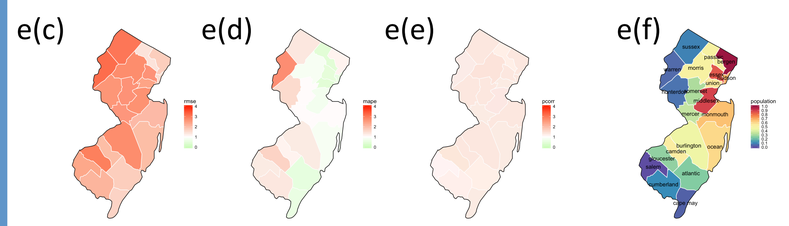
$$\min_{\theta} \mathcal{L}(\theta) = \sum_t \|\mathbf{z}_t - f([\mathbf{x1}, \mathbf{x2}]_t, \theta)\|_2^2 + \mu\phi(\hat{\mathbf{z}}_t) + \lambda\delta(\hat{\mathbf{z}}_t),$$

$$\phi(\hat{\mathbf{z}}_t) = \left|\hat{y}_t - \sum_{C \in \mathcal{D}} \hat{y}_t^C\right|, \qquad \delta(\hat{\mathbf{z}}_t) = \left|\frac{1}{K+1}\sum \max(-\hat{\mathbf{z}}_t, \mathbf{0})\right|,$$

spatial consistency **non-negative consistency**

Experiment Results





e(a) – e(b) State-level and county-level forecasting performance over two states (VA and NJ) and two seasons. e(c) – e(e) TDEFSI over EpiFast on NJ county-level forecasting performance by RMSE, MAPE and PCORR. e(f) Population heat map of NJ counties.

Highlights TDEFSI 1. trains on theory generated synthetic data; 2. integrates the strengths of DNNs and high-resolution simulations of epidemic processes over networks; 3. yields accurate high-resolution spatiotemporal forecasts using low-resolution time series data.