ON THE PREDICTABILITY OF SPATIOTEMPORAL EPIDEMICS

Jeremy Nachison, Dr. Srini Venkatramanan

Motivation

Scarpino and Petri's seminal paper "On the Predictability of Infectious Disease Outbreaks"[1] claims that single outbreaks should be temporally predictable. We adapt methods from medical imaging and neuroscience to extend Scarpino's analysis into the spatiotemporal domain. Using a metric for the predictability of an epidemic across both space and time, we investigate how spatial information influences epidemic predictability:

- Can spatial information make epidemics more predictable and allow for longer forecast horizons?
- Does accounting for additional complexity in the system increase predictability?
- How does predictability vary with disease parameters and network connectivity for synthetic epidemics?
- How do predictability metrics align with the general understanding of COVID-19 and seasonal flu waves?

Background

Using Permutation Entropy[2] (PE) across timeseries data for multiple infectious diseases as a measurement of predictability, Scarpino et al. makes the following conclusions:
- Differences in the average reproductive number and contact network structure can drive differences in predictability
- As outbreaks become entangled with human mobility, pathogen evolution etc. they become increasingly less predictable

PE has been adapted to many different applications and extended to measure complexity on complex data:
- Schlemmer et al.[3] adapts PE to measure complexity across cardiac arrythmia video data via spatiotemporal image embeddings.
- Boaretto et al.[4] uses PE with spatial embeddings to distinguish resting brain states and show temporal embeddings alone fail to do so.
- Fabila-Carrasco[5] adapts PE to measure complexity of data along graphs.

Methods

We adapt the spatiotemporal embedding formulations of (3) and (4) to suit timeseries data across and define Spatiotemporal Epidemic Permutation Entropy (STEpPE):

Let M be a matrix that holds the data for location j at timestep i. We construct a temporal embedding vector $x_{ij}^{d_t,\tau}$ for entry M_{ij} with dimension d_t, and temporal separation τ and a spatial embedding vector $x_{ij}^{d_s,\omega}$ with dimension d_s and spatial separation ω. Define the temporal embedding vector as:

$$x_{ij}^{d_t,\tau} = [M_{i,j}, M_{\tau+i,j}, \dots, M_{(d_t-1)\tau+i,j}]$$

Let $d(j_1, j_2)$ denote the distance between locations j_1 and j_2. Then,

$$x_{ij}^{d_s,\omega} = [M_{i,j}, M_{i,j_\omega}, \dots, M_{i,j_{(d_s-1)\omega}}]$$

Where $\forall k < l, d(j, j_k) \le d(j, j_l)$. I.e. form a vector of the $d_s - 1$ nearest neighbors of location j. The spatiotemporal embedding vector of dimension $\mathbf{d} = d_t + d_s - 1$ is then the union of these embeddings:

$$x_{ij}^{d} = [x_{ij}^{d_t,\tau} \cup x_{ij}^{d_s,\omega}]$$

By ranking the entries of x_{ij}^{d} it is assigned to one of the $d!$ permutations of orderings of d integers π_k. Let p_{π_k} denote the percentage of embedding vectors assigned to π_k. We then compute the STEpPE as:

$$STEpPE(M_j) = -\frac{1}{\log(d!)} \sum_{\pi_k} p_{\pi_k} \log(p_{\pi_k})$$

The predictability is the difference between the entropy which would result from uniformly distributed ordinal patterns, 1, and the entropy of the ordinal data distribution: $1 - STEpPE(M_j)$.

We analyze how the predictability of simulated metapopulation epidemics is affected by the embedding dimension, importation level, exposure and recovery rates, reshuffling of network edges, and adding realistic surveillance noise. We then compare COVID-19 and Influenza hospitalizations and examine how timeseries length and reshuffling of distances between states impacts STEpPE and PE.

Results

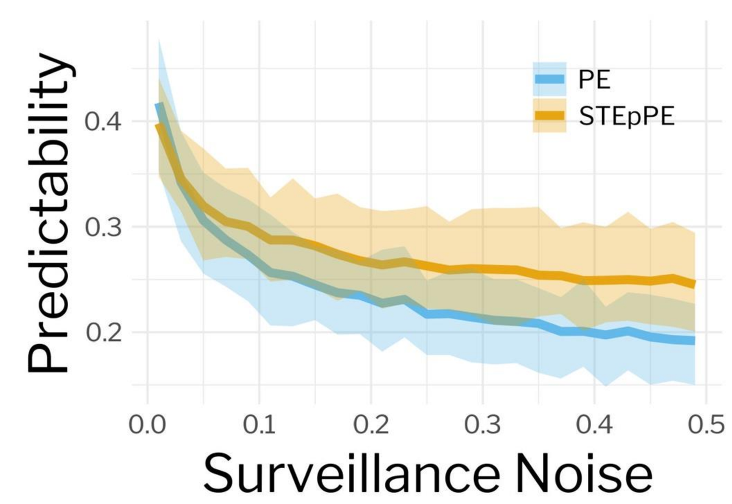

Figure 1) Predictability by PE and STEpPE through increasing surveillance noise.

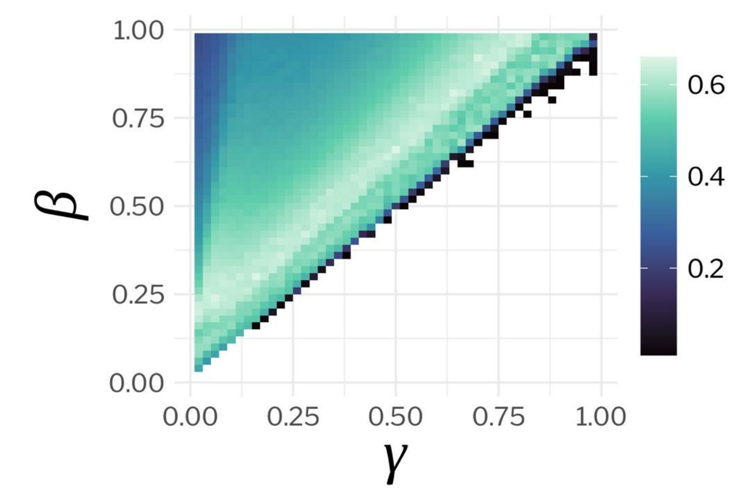

Figure 2) Contour of STEpPE as a function of exposure rate, β, and recovery rate, γ.

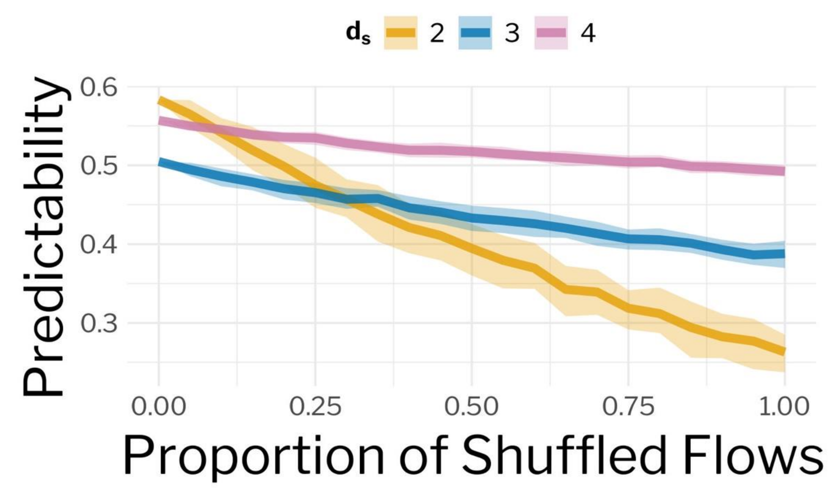

Figure 3) Predictability of simulated epidemic vs. shuffling of network edges.

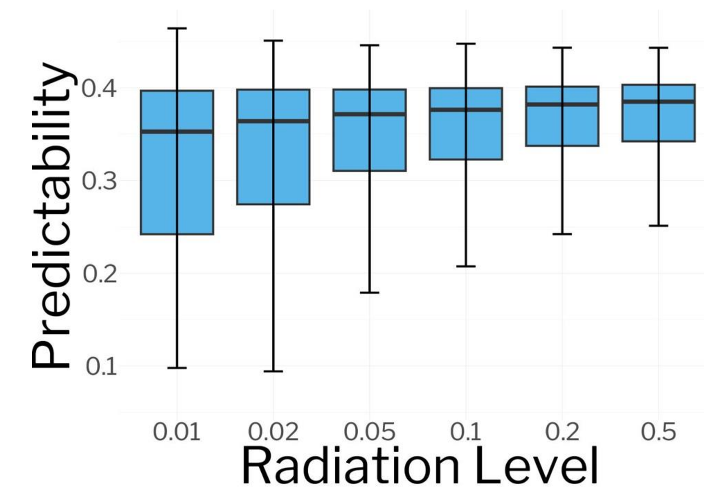

Figure 4) Predictability through network connectivity levels.

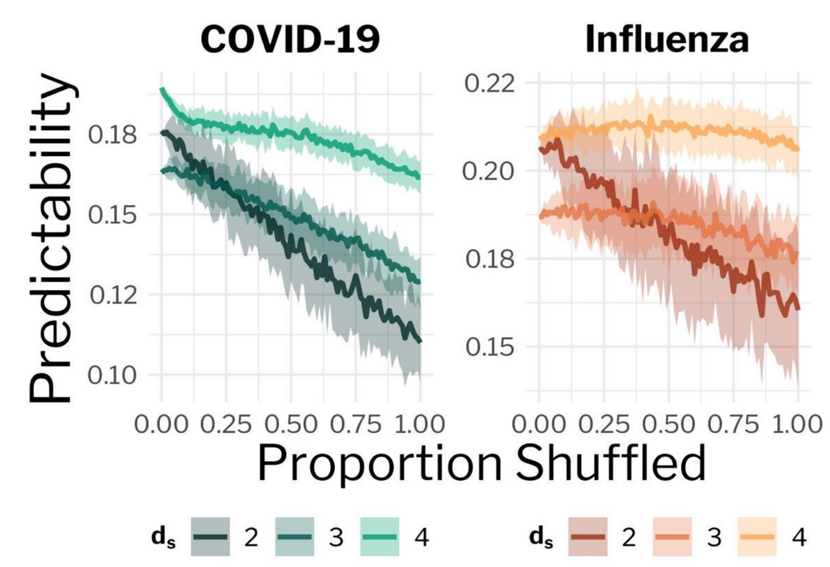

Figure 5) Predictability of COVID-19 and Influenza vs. shuffling of state location data.

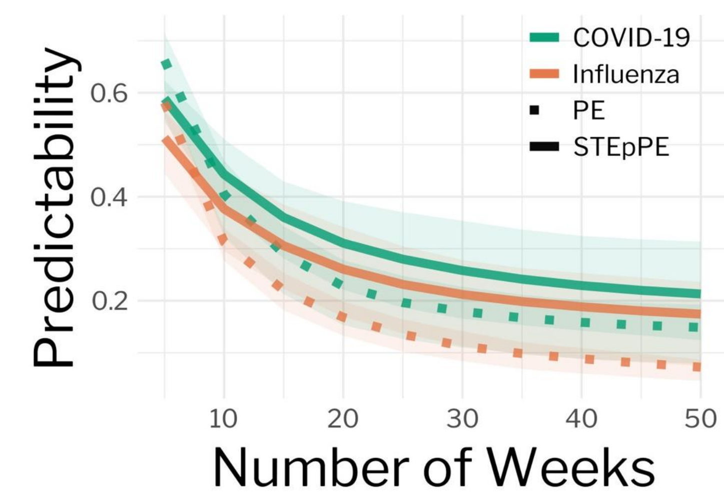

Figure 6) Predictability by PE and STEpPE for Influenza and COVID-19 with increasing timeseries length.

Discussion

- Predictability under STEpPE decreases at a slower rate compared to PE, allowing for higher predictability at longer timeseries lengths, but may be less predictable in the short term.
  - The incorporation of spatial information increases predictability of Influenza more than COVID-19.
  - Influenza is more predictable across space.
  - COVID-19 and Influenza are sensitive to spatial perturbations, albeit less than the simulated epidemic.
- As the flow between locations increases, the predictability of the epidemic becomes less sensitive to the pre-epidemic phase.
- The most unpredictable epidemics are those with R_0 only slightly above the critical threshold, and diseases with the highest epidemic potential are more predictable.

Acknowledgements

This work was partially funded by NSF Expeditions in Computing, CDC/CSTE FluSight, NIH MIDAS program, and UVA Strategic Investment Fund.

References

1. Scarpino, S. V., & Petri, G. (2019). On the predictability of infectious disease outbreaks. Nature communications, 10(1), 898.
2. Bandt, C., & Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. Physical review letters, 88(17), 174102.
3. Schlemmer, A. (2018). Spatiotemporal permutation entropy as a measure for complexity of cardiac arrhythmia. Frontiers in Physics, 6, 39.
4. Boaretto, B. (2023). Spatial permutation entropy distinguishes resting brain states. Chaos, Solitons & Fractals, 171, 113453.
5. Fabila-Carrasco, J. S., Tan, C., & Escudero, J. (2022). Permutation entropy for graph signals. IEEE Transactions on Signal and Information Processing over Networks, 8, 288-300.

University of Virginia
BIOCOMPLEXITY INSTITUTE