History Repeats Itself: COVID-19 Forecasting

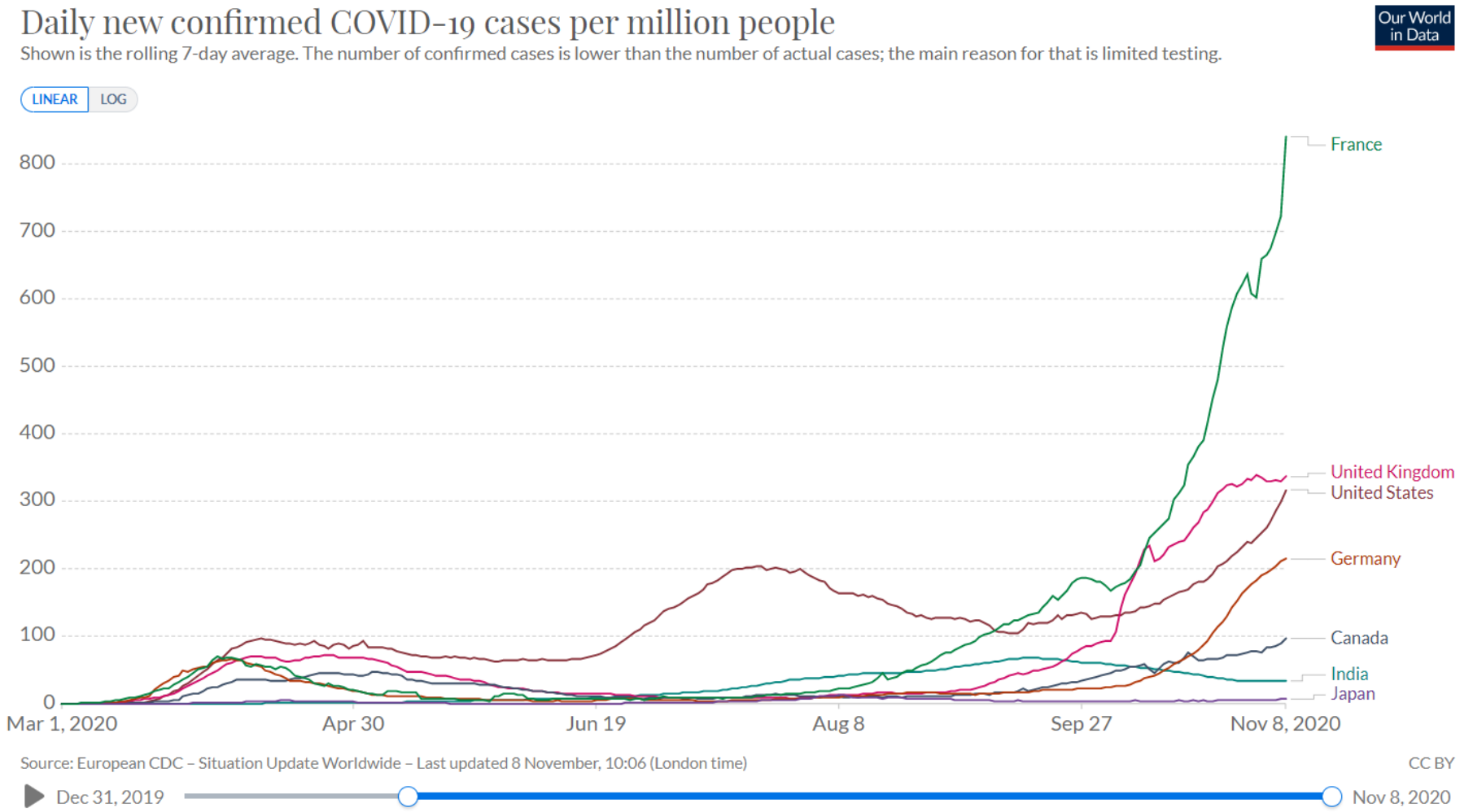
Prof. Xifeng Yan

Collaborators: Xiaoyong Jin, Prof. Yu-Xiang Wang

Computer Science, UCSB

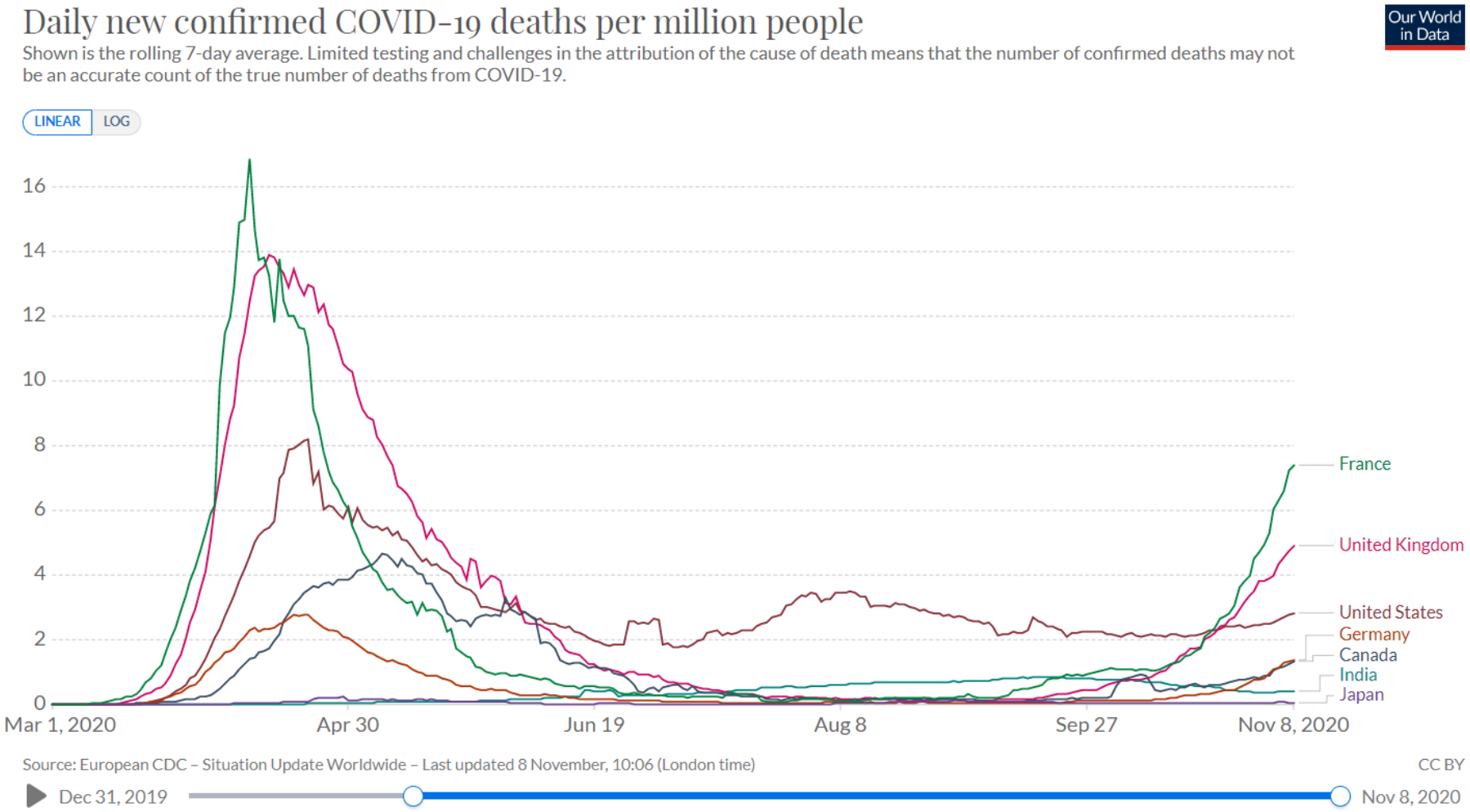
NSF Rapid: Interventional COVID-19 Response Forecasting

Daily New Confirmed Cases



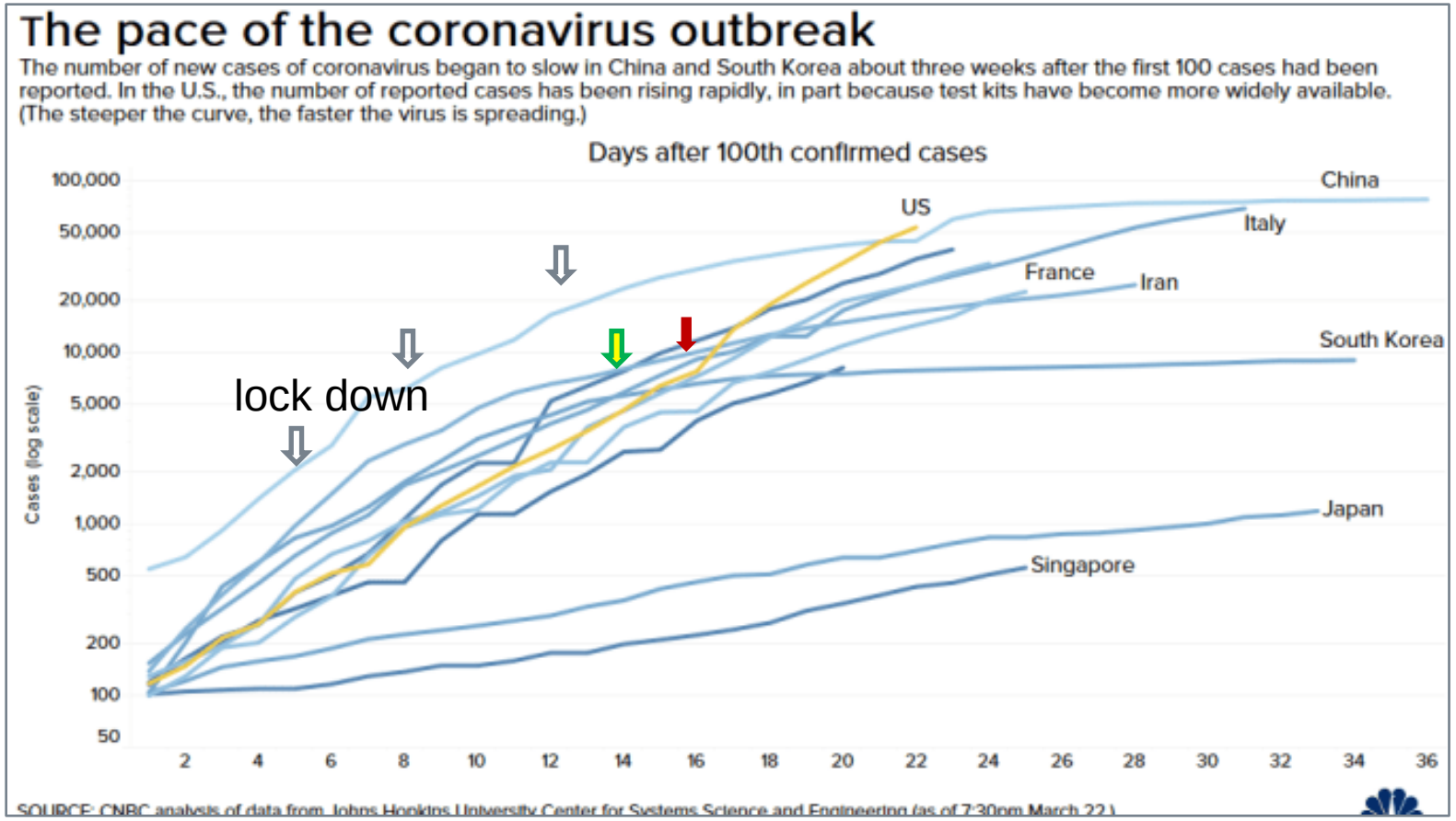
credit: https://ourworldindata.org/coronavirus

Daily New Deaths



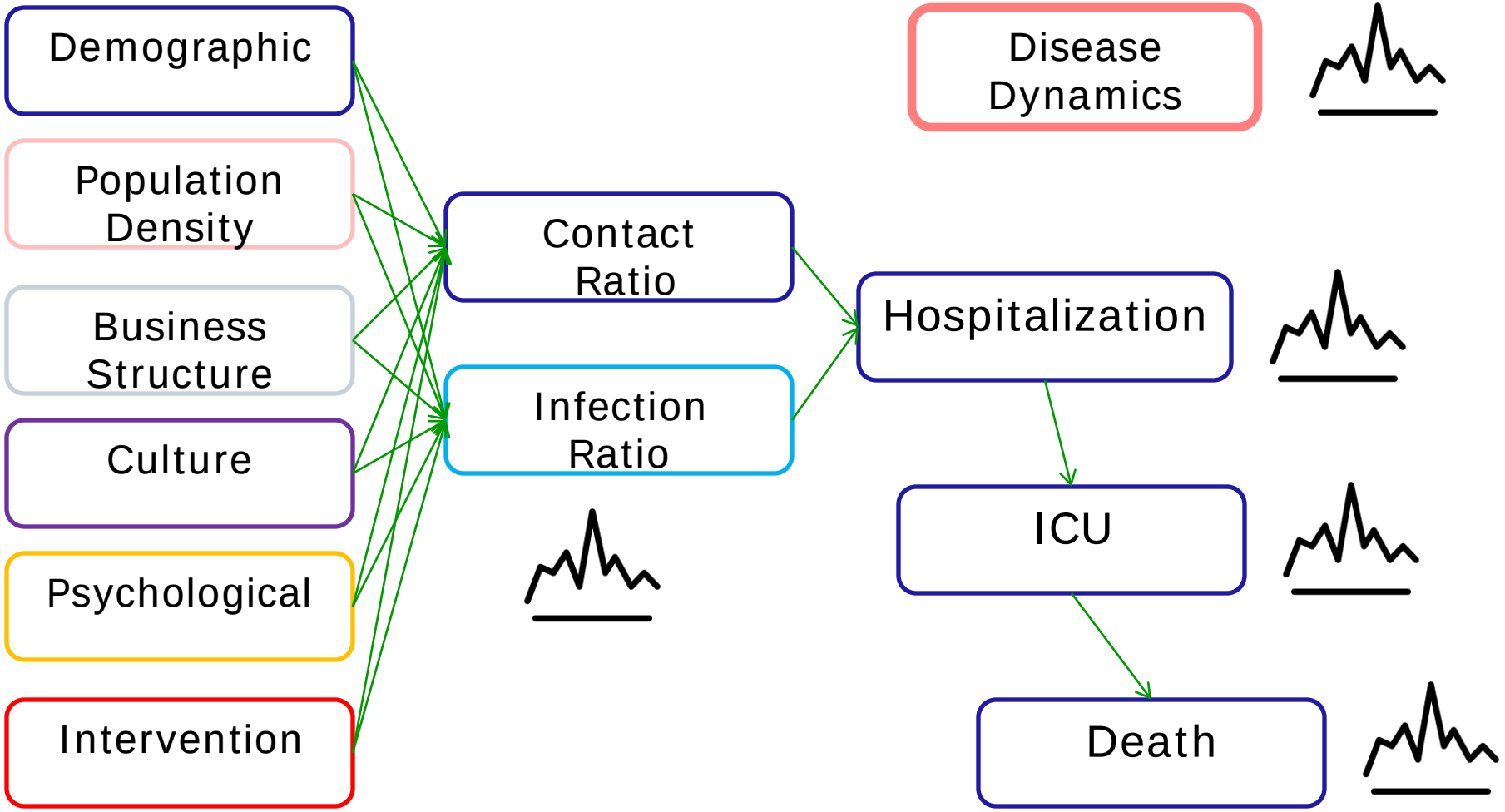
credit: https://ourworldindata.org/coronavirus

NSF Rapid: Interventional COVID-19 Response Forecasting

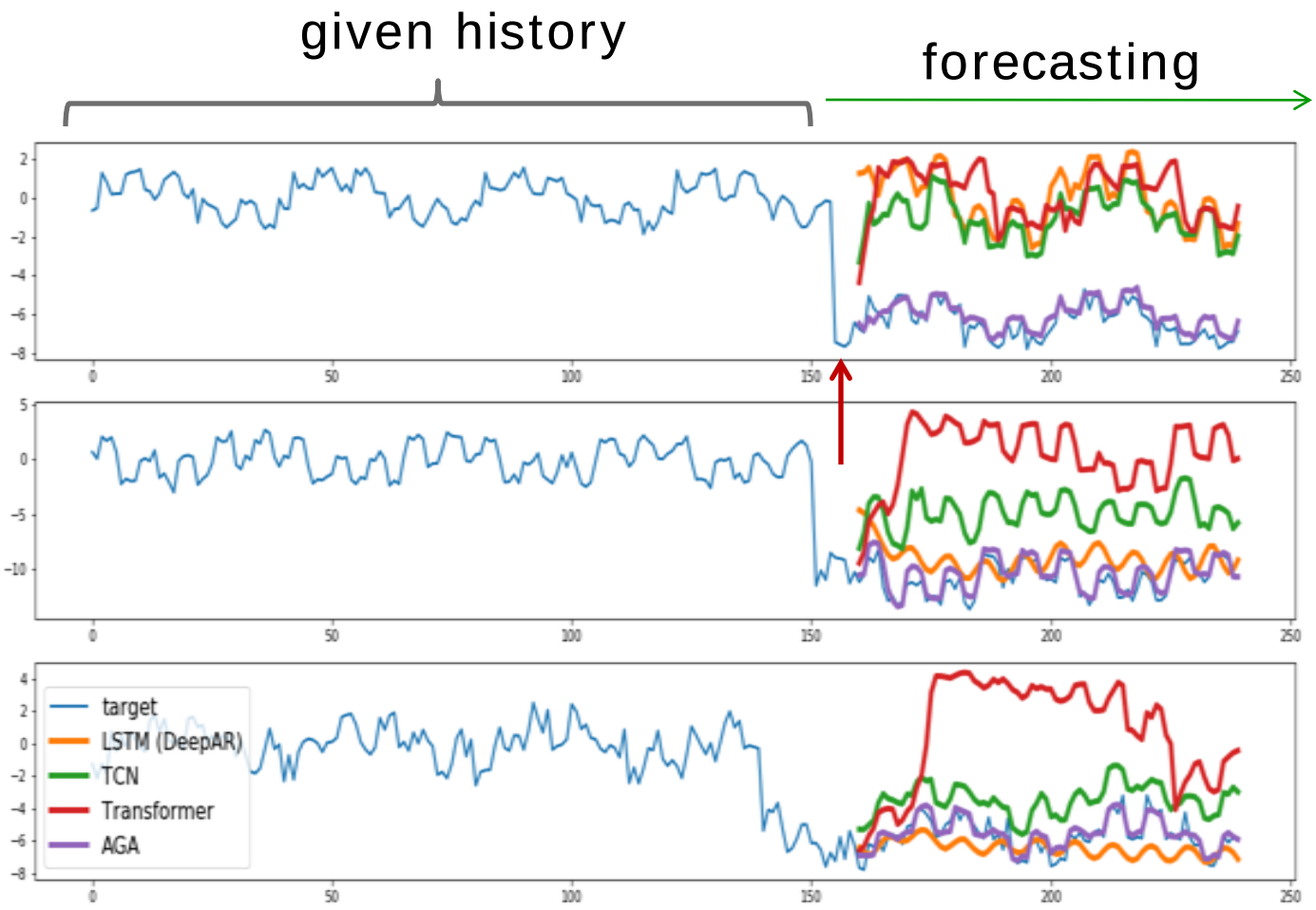


We started the project in March, 2020

Many Factors: Some are Very Dynamic



Time Series Forecasting: Extensively Studied



Classic Epidemic Models: SIR, SEIR etc.

- Modeling Spread of Disease

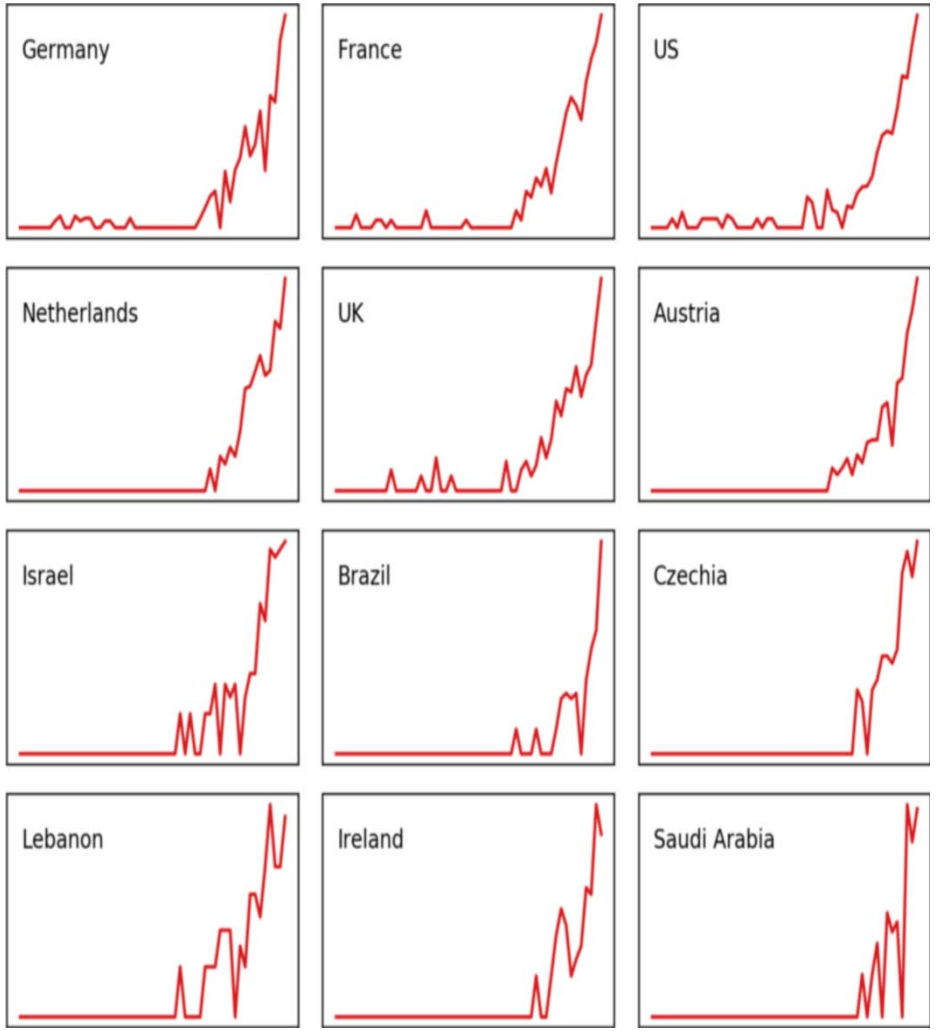
$$\frac{dS}{dt} = -\frac{\beta I S}{N},$$

$$\frac{dI}{dt} = \frac{\beta I S}{N} - \gamma I,$$

$$\frac{dR}{dt} = \gamma I$$

Try Something Different: History Repeats Itself

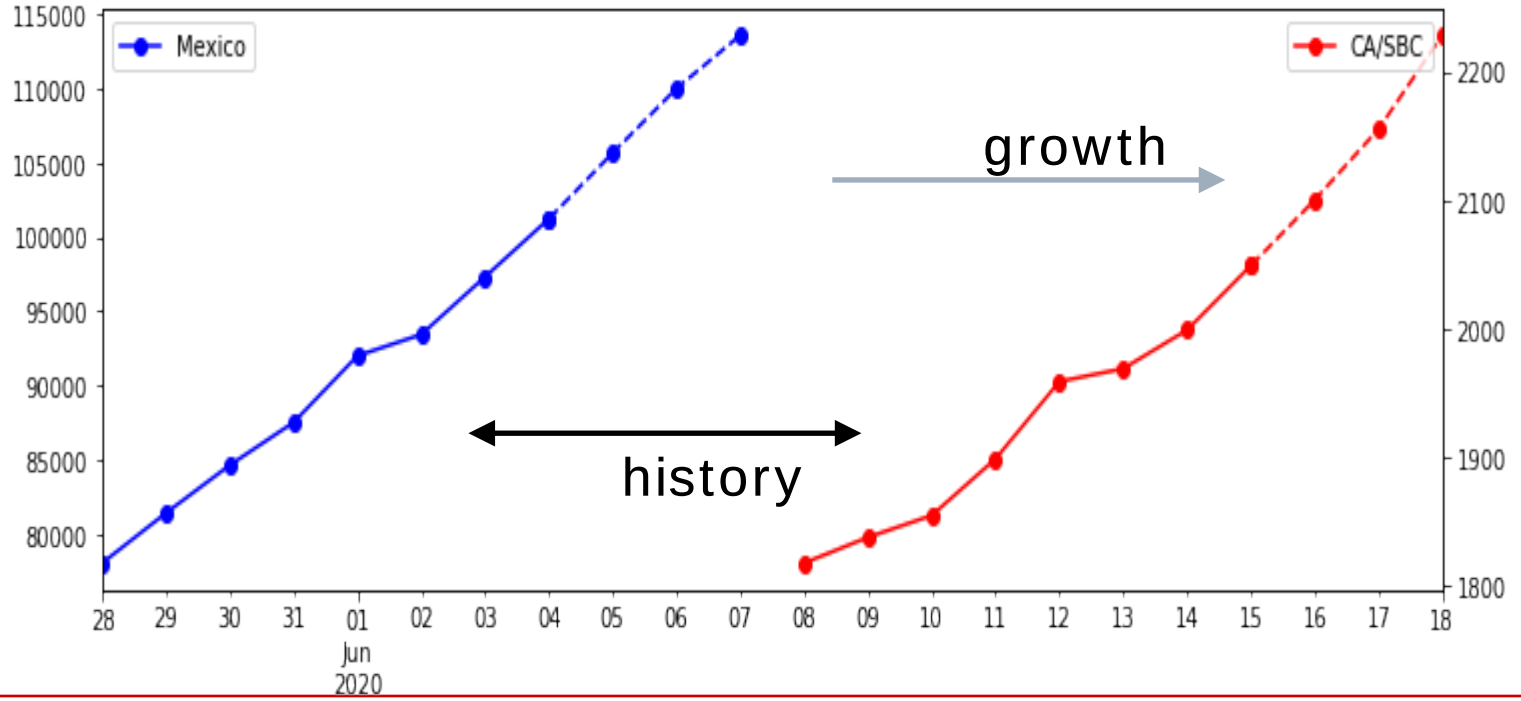
- Different regions share COVID-19 trending pattern
  - The spreading rate is determined by common factors, such as social interactions and protections
- To forecast cases in a certain region, we can refer to other regions where pandemic starts earlier



Daily new cases

Similar historical pattern → Similar future growth

- In July, Santa Barbara county was experiencing a new wave of COVID-19 spreading that resembles that in other regions
  - e.g. Mexico in early June



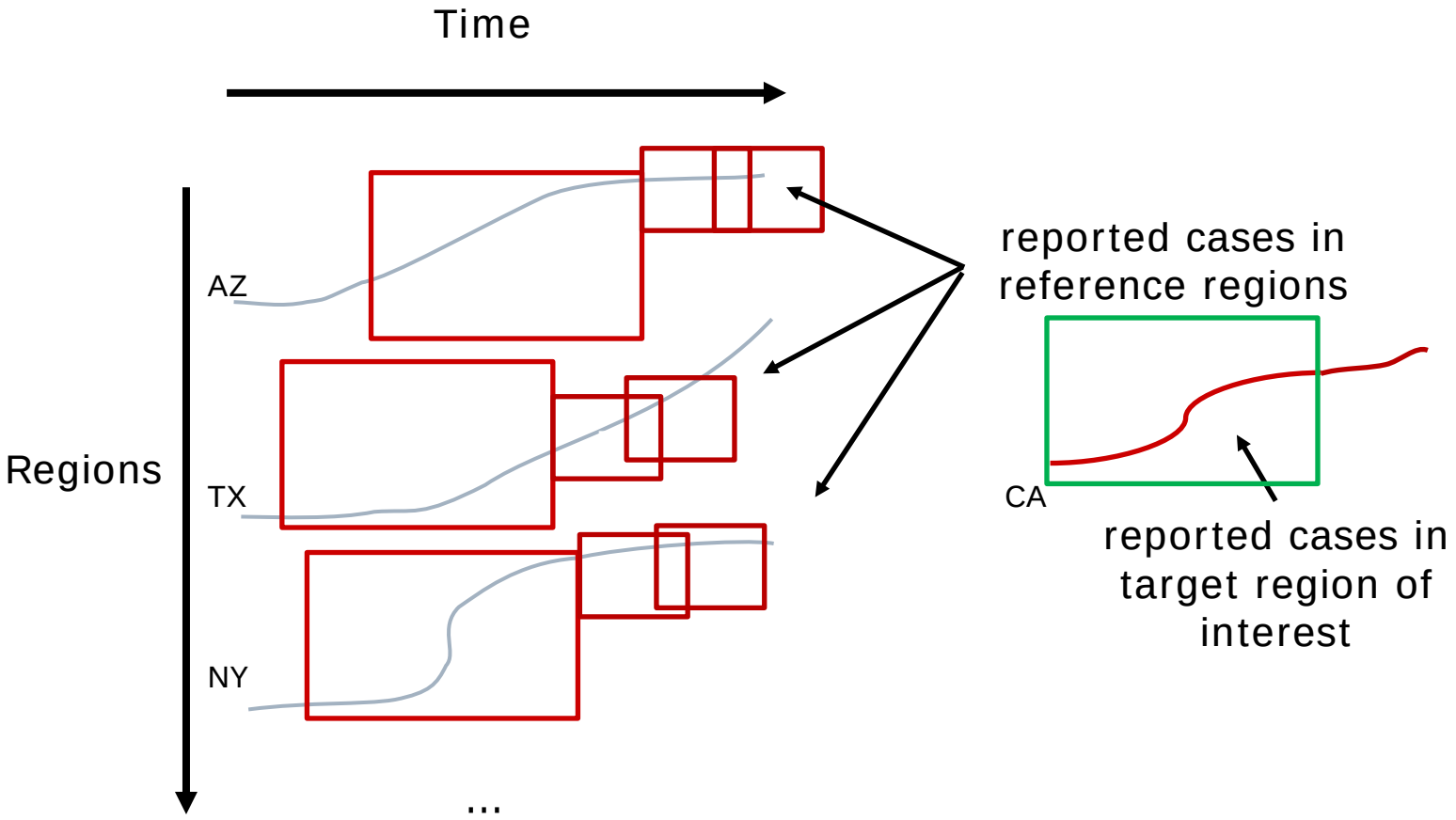
Find Similar Regions (Reference)

- For example, sharing similar factors : demographic, population density, business structure, social culture, psychological factors and interventional policy

Or straightforward

- Find regions whose trends look similar: All the aforementioned factors have been priced in!

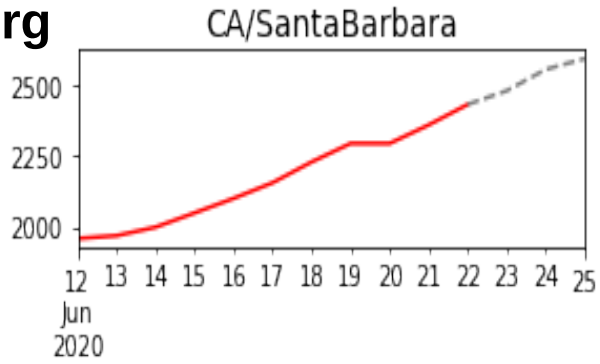
Forecasting with Attention Mechanism

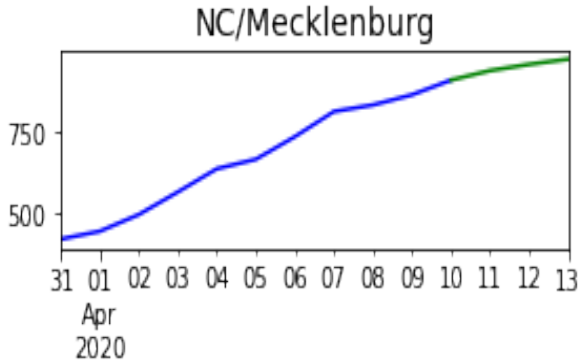


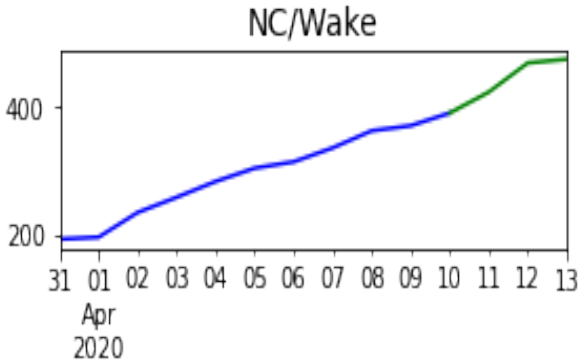
Identified References

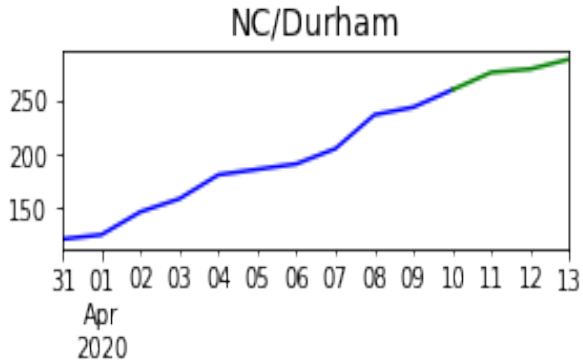
For Santa Barbara County during 06/12 ~ 06/22

- Most similar state: **North Carolina** during 03/31~04/10
- Most similar counties in NC:
  - NC/**Mecklenburg**
  - NC/**Wake**
  - NC/**Durham**

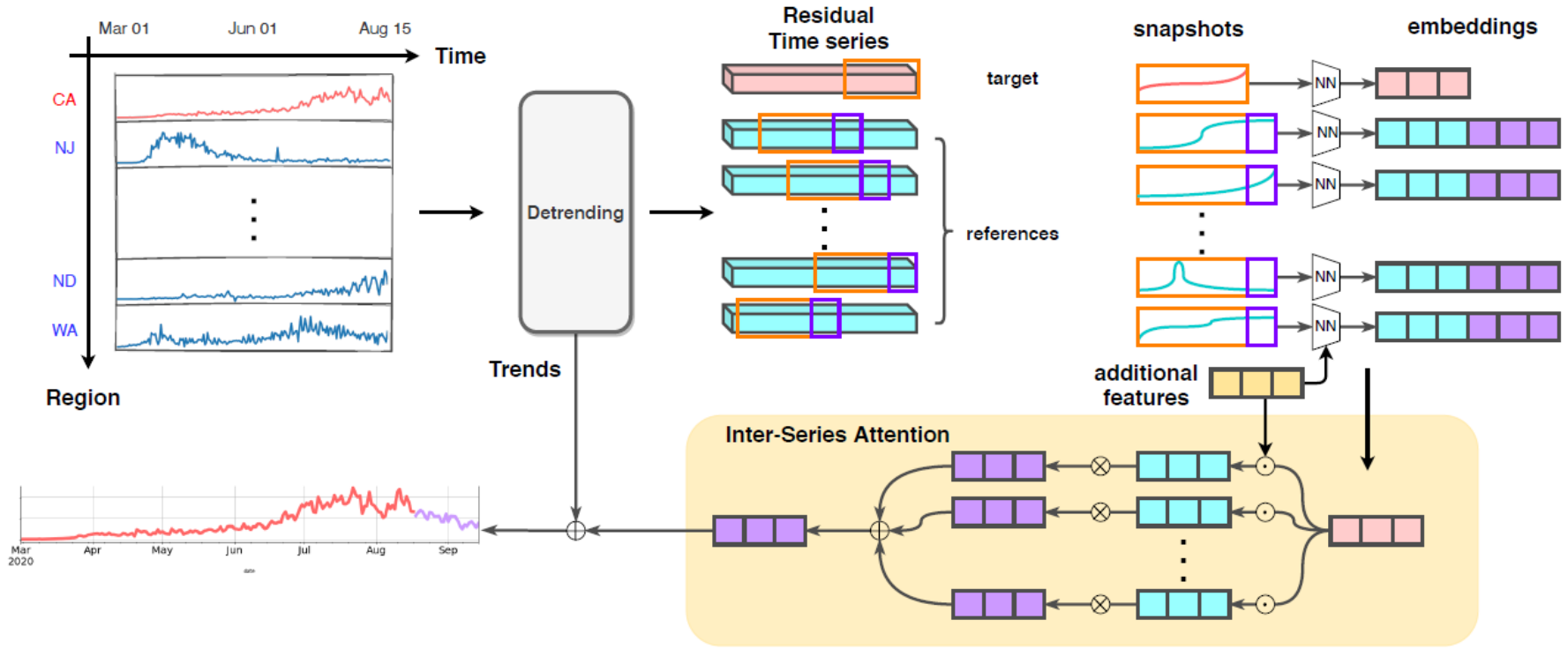








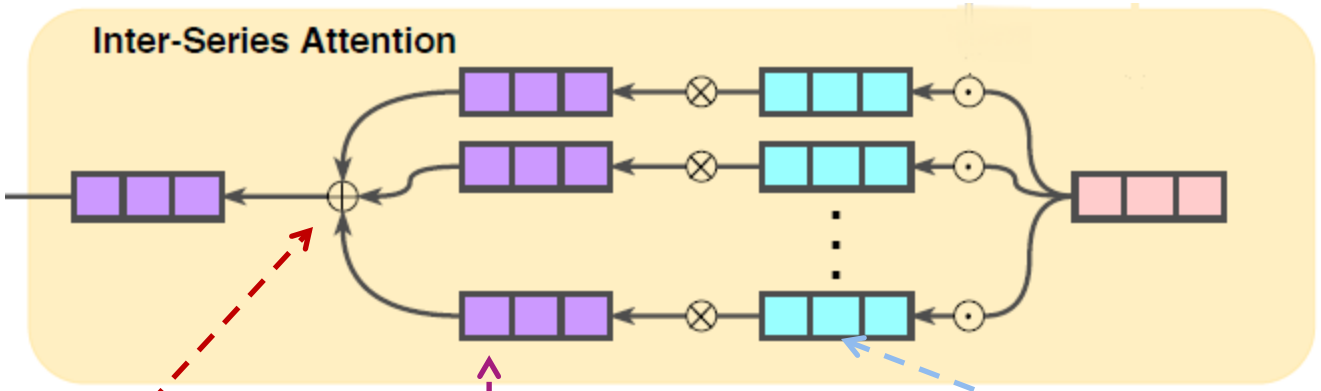
Inter-Series Forecasting Model (ACTS)



an end-to-end training process to minimize the total error

Paper link: https://arxiv.org/abs/2010.13006

Inside ACTS



Inter-series Attention

$$\boldsymbol{q}_t^i = \boldsymbol{W}_Q \boldsymbol{p}_t^i + \boldsymbol{W}_{u,q} \boldsymbol{u}^i;$$
$$\boldsymbol{k}_t^i = \boldsymbol{W}_K \boldsymbol{p}_t^i + \boldsymbol{W}_{u,k} \boldsymbol{u}^i;$$
$$\boldsymbol{v}_t^i = \boldsymbol{W}_V \boldsymbol{g}_t^i;$$

$$\hat{\boldsymbol{v}}_T^{i_0} = \sum_{i,t \in \Omega} \frac{\exp\left(\langle \boldsymbol{q}_T^{i_0}, \boldsymbol{k}_t^i \rangle\right)}{\sum_{i,t \in \Omega} \exp\left(\langle \boldsymbol{q}_T^{i_0}, \boldsymbol{k}_t^i \rangle\right)} \boldsymbol{v}_t^i$$

time series embedding

$$\boldsymbol{p}_t^i = \mathrm{AvgPool}\left(\mathrm{Conv}\left(\left[\tilde{\boldsymbol{c}}_{t-l+1:t}^i; \boldsymbol{r}_{t-l+1:t}^i\right]\right)\right)$$

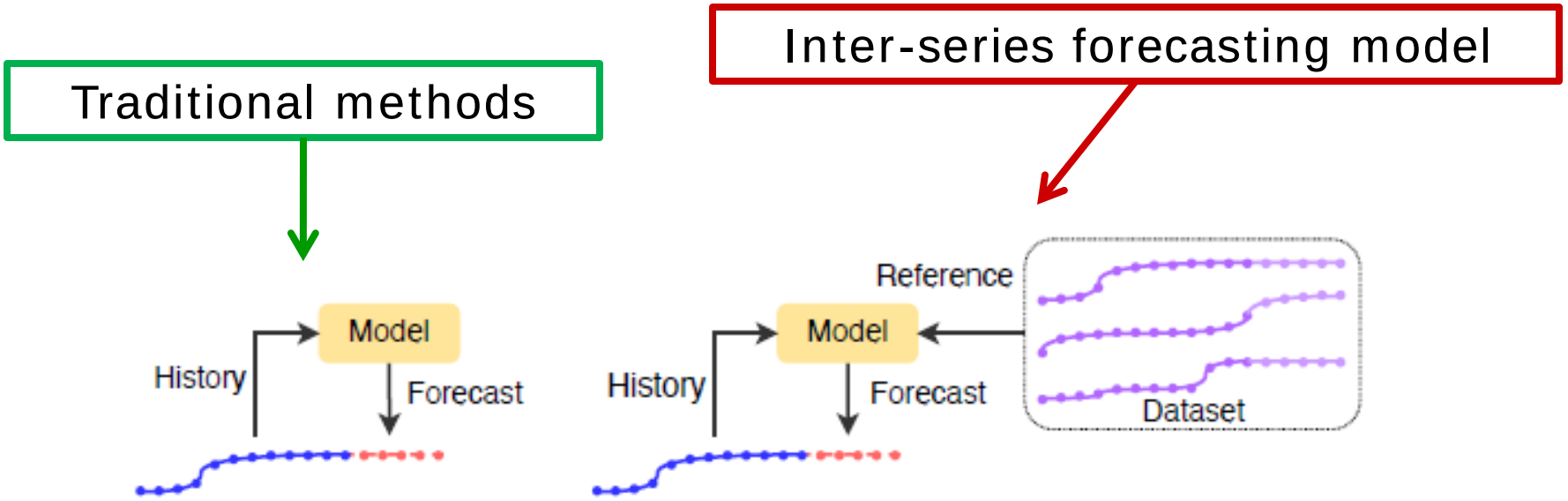
note: not directly compare the curve shape

follow-up trend embedding

$$\boldsymbol{g}_t^i = \mathrm{AvgPool}\left(\mathrm{Conv}\left(\tilde{\boldsymbol{c}}_{t+1:t+H}^i\right)\right)$$

note: not directly use the curve value

Another View



CDC: *COVID-19 Forecasting*

		Method						measure:WAPE			ours
		YYG	CU	UCLA	ERDC	LANL	Covid Sim	Deep COVID	Conv Trans	TFT	ACTS
06/21	C	-	-	-	-	0.51	-	-	1.09	0.51	**0.39±0.01**
	H	-	1.91	-	-	1.08	0.95	**0.63**	1.22	0.80	0.80±0.02
	D	0.52	1.48	0.56	-	0.58	1.46	0.66	1.09	0.67	**0.45±0.01**
07/05	C	-	-	-	-	0.37	-	-	0.37	0.39	**0.33±0.01**
	H	-	0.98	1.23	0.66	0.95	-	0.65	1.08	0.84	**0.61±0.04**
	D	0.45	0.65	0.53	**0.38**	0.52	-	0.85	0.60	0.51	0.60±0.01
07/19	C	-	-	-	-	**0.27**	-	-	0.50	0.44	0.31±0.01
	H	-	0.67	1.24	0.77	0.78	1.71	0.70	0.99	0.66	**0.60±0.03**
	D	0.30	0.43	0.39	1.10	0.48	0.33	0.4506	0.54	0.67	**0.28±0.01**
08/02	C	-	-	-	-	0.30	-	-	0.24	0.24	**0.16±0.04**
	H	-	0.67	0.95	0.71	0.68	1.66	0.79	0.93	0.92	**0.66±0.09**
	D	0.24	0.37	0.27	0.57	0.44	0.26	0.29	0.45	0.38	**0.21±0.01**
08/16	C	-	0.67	0.35	0.28	0.29	0.23	-	0.33	0.55	**0.20±0.03**
	H	-	0.64	0.99	0.60	0.65	1.38	0.98	0.96	0.92	**0.57±0.02**
	D	**0.19**	0.42	0.25	0.53	0.34	0.27	0.28	0.44	0.31	0.23±0.01
08/30	C	-	0.43	0.31	0.34	0.33	0.23	-	0.36	0.29	**0.23±0.03**
	H	-	0.66	0.91	0.68	0.69	1.31	0.83	0.93	0.82	**0.58±0.03**
	D	**0.20**	0.41	0.23	0.56	0.34	0.25	0.36	0.42	0.40	0.25±0.02

C: New Cases, H: Hospitalizations, D: Deaths

https://github.com/Gandor26/covid-open

Many Models Are Improving

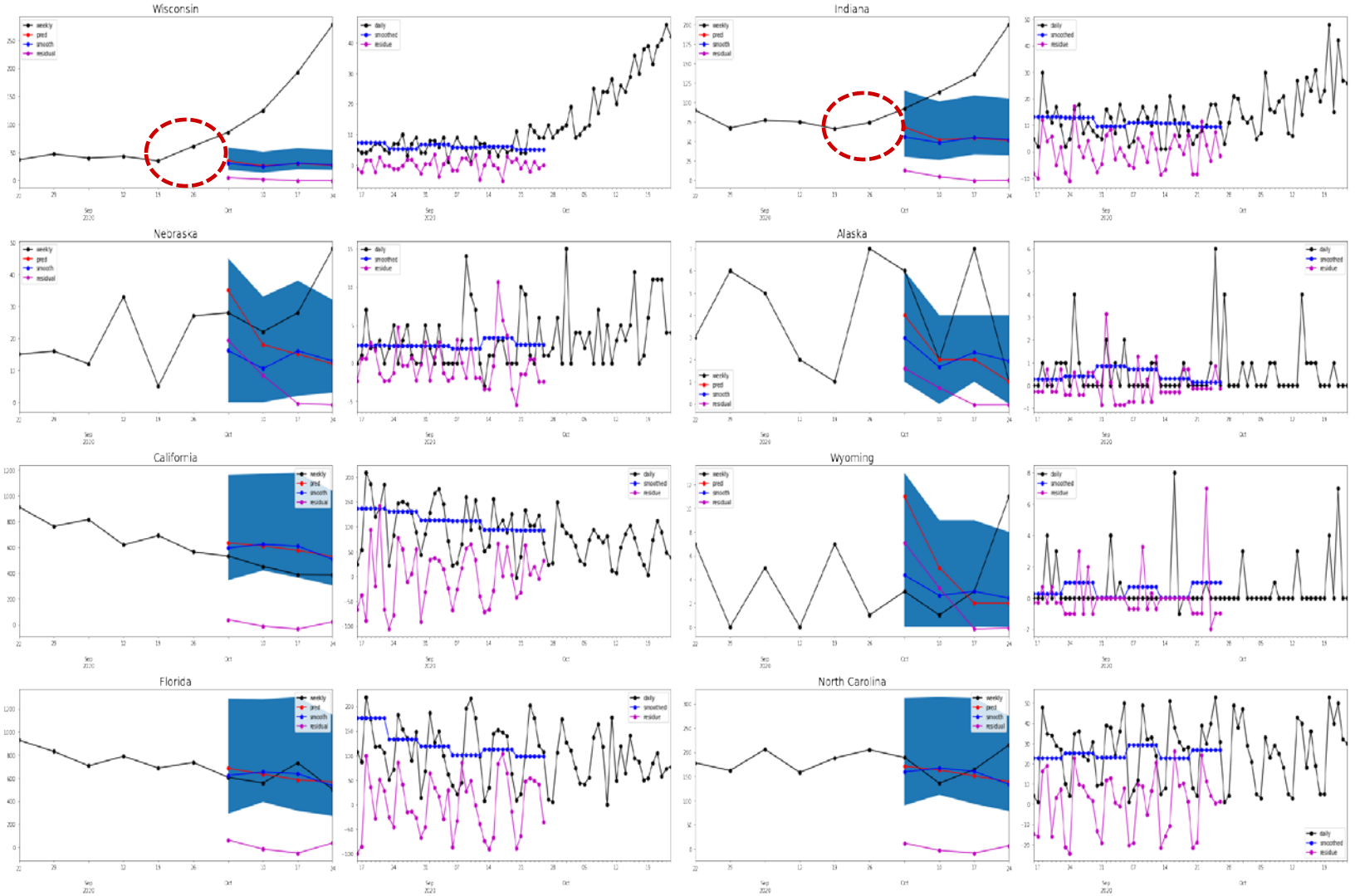
ours

		YYG	CU	UCLA	ERDC	COVID Sim	Deep COVID	LANL	ACTS
09-14	C	-	0.43	0.40	0.37	-	-	0.42	**0.34**
	H	-	0.74	0.69	0.66	-	0.67		**0.64**
	D	0.28	0.40	0.26	0.58	-	**0.27**	0.37	0.36
09-28	C	-	**0.35**	0.41	**0.35**	**0.35**	-	0.47	0.46
	H	-	0.65	0.71	0.69	0.81	0.65	0.75	**0.63**
	D	**0.27**	0.49	0.32	0.64	0.30	0.28	0.40	0.34
10-12	C	-	**0.33**	0.56	0.37	0.43	-	0.55	0.45
	H	-	0.61	0.78	0.69	0.90	0.60	0.61	**0.57**
	D	-	**0.31**	0.42	0.69	0.34	0.32	0.46	0.41

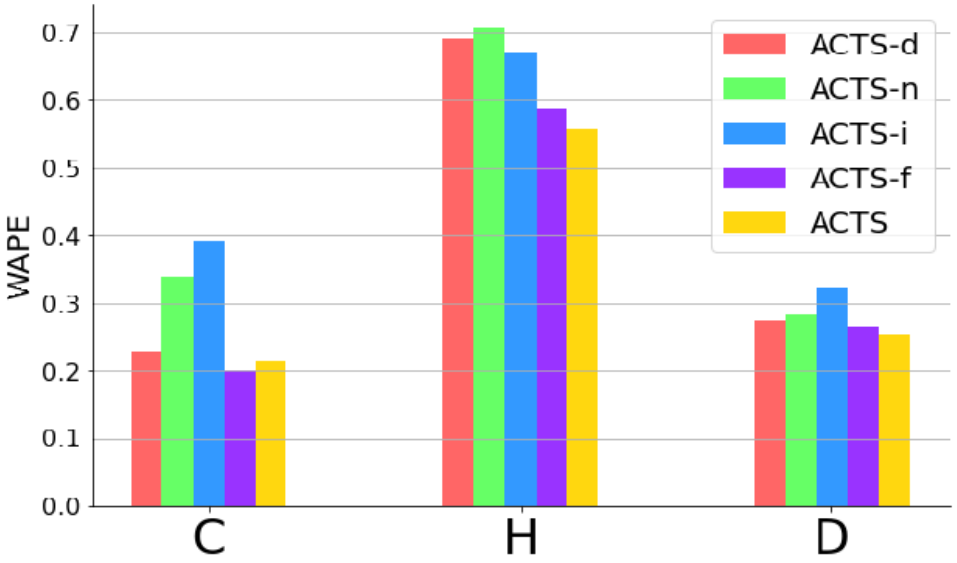
https://github.com/Gandor26/covid-open

warning: It doesn't indicate future performance

Result Analysis : Deaths



Ablation Study



C: New Cases, H: Hospitalizations, D: Deaths

- ACTS-d: Remove the detrending module
- ACTS-n: Remove the normalization in segment embedding
- ACTS-i: Restrict the attention to the target time series only.
- ACTS-f: Remove the additional features

Claim: If we directly use the Method of Analogues (Viboud et al. 2003), the error will be much higher as it is a model closer to ACTS-d-n-i-f .

Future Developments

Data-Driven Neural Models

+ SEIR

+ Global Trend

+ Additional Data

Currently we only use cases, hospitalizations, deaths, some limited demographical information. If you have other data sources or links, please let us know.

Q&A

Xifeng Yan

xyan@cs.ucsb.edu