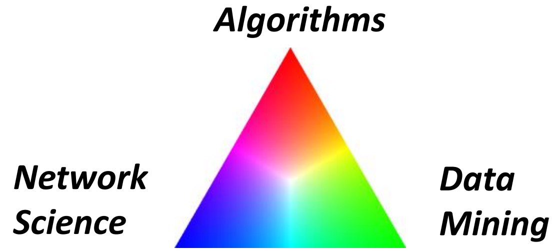
Learning to Forecast and Forecasting to Learn from the COVID-19 Pandemic

Ajitesh Srivastava, Viktor Prasanna
Data Science Lab

University of Southern California

Data-driven Research

My work at Data Science Lab
dslab.usc.edu
Led by Prof. Viktor K. Prasanna



Abstract
Mathematical formulation, provably correct solutions

Experimental
Justified intuition for methods from assumptions, experimentally verifiable solution

Real World
Justified intuition from observations/domain experts, practical, deployable solution

GraphSAINT: State-of-the-art Embedding

Influence Computation, Maximization, Competing Cascades

Restricting FakeNews

High-throughput FPGA solutions

Image classification

Load forecasting

Crime forecasting

ML-driven prefetcher

ML-driven compiler

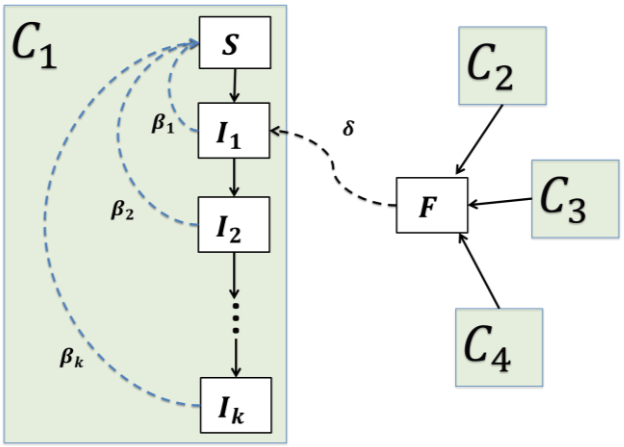
Target Challenge: Predicting what shoppers will buy

Violence reduction among homeless

DARPA Challenge Chikungunya Epidemic Forecasting

COVID-19: Forecasts, Resource allocation, Restarting the economy, …

DARPA Grand Challenge – CHIKV (2014-2015)



Heterogeneous infection rate model with human mobility

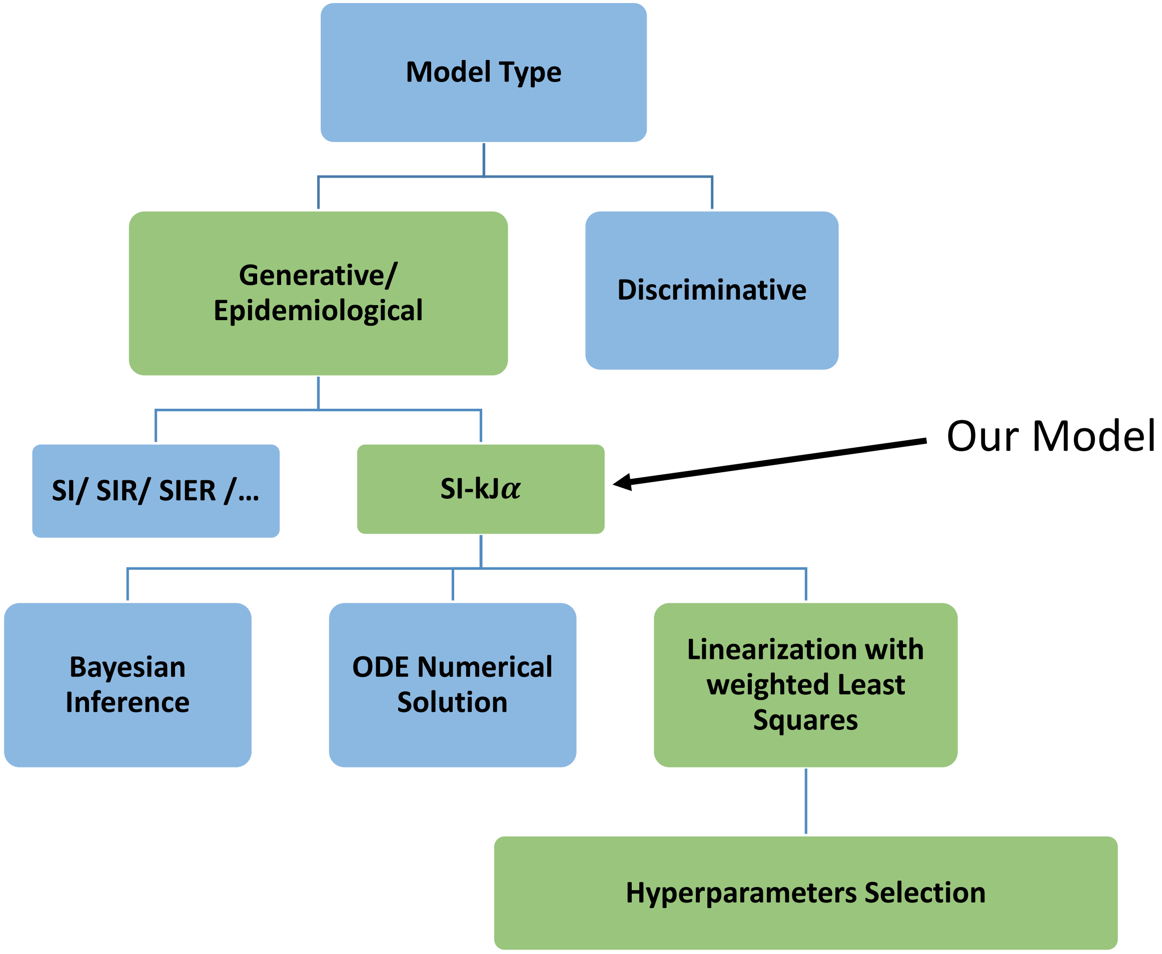
CHIKV epidemic: Country-level predictions. Weekly over 8 months, 55 countries

One of 10 winners of DARPA Grand Challenge 2015 for predicting CHIKV epidemic

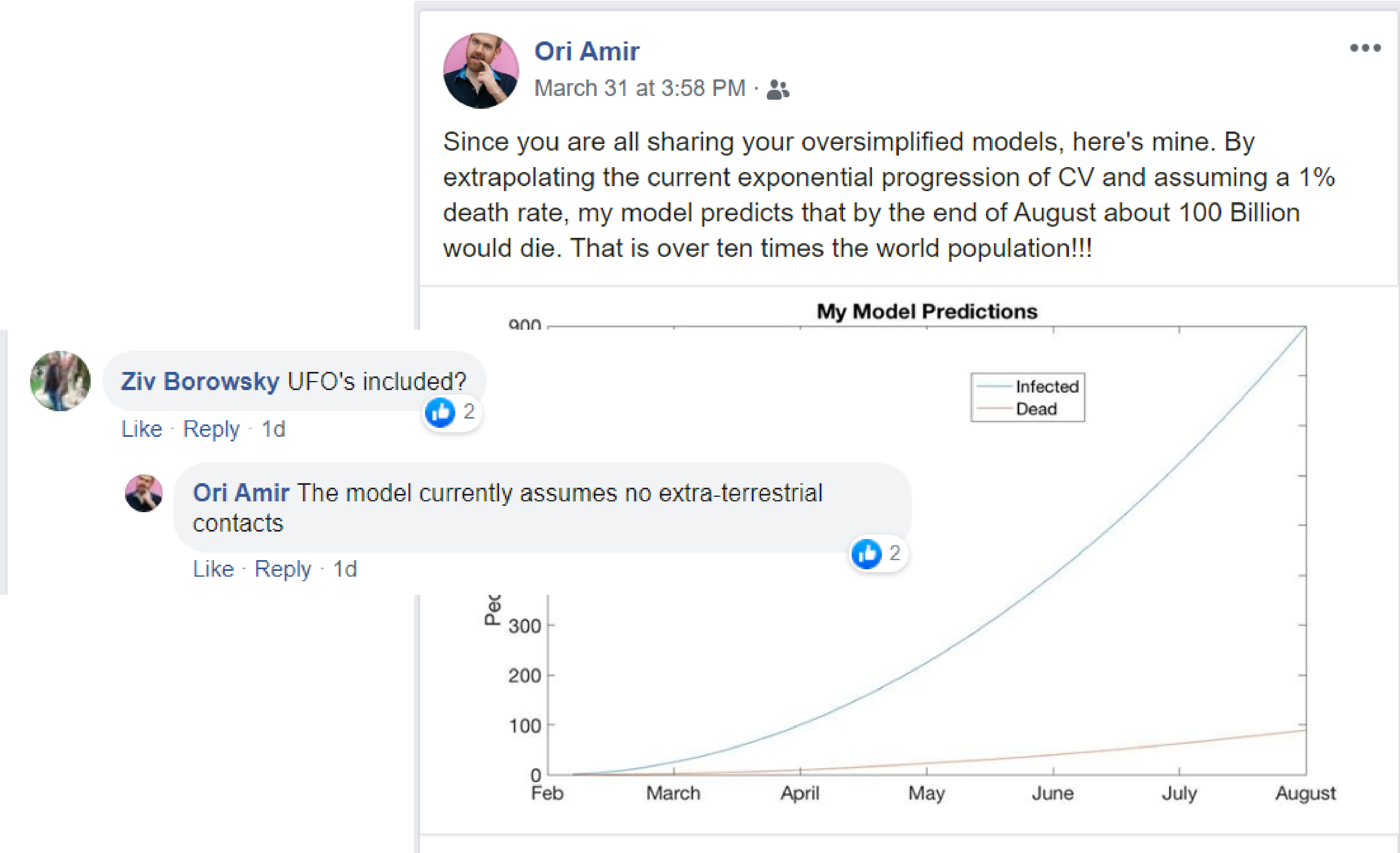
Why Forecast?

- Preparedness and resource management needs **state/county/city level predictions**:
  - How many masks, testing kits, beds are needed tomorrow/next week at a given hospital
  - How to distribute state/country resources across all the hospitals in a state/country
- How do we come out of "stay-at-home" order?
  - Should some venues remain closed and some open, initially?
- Need accurate forecasts and simulations of future scenarios

Modeling Choices



Forecasting is Difficult

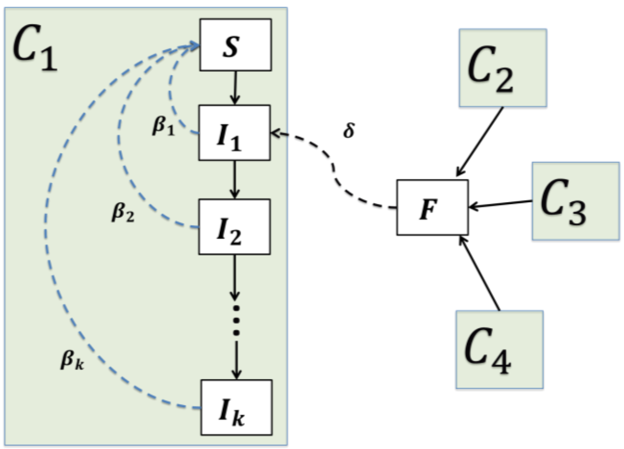


SI-kJα - Heterogeneous Infection Rate with Human Mobility

$$\Delta I_t^p = \frac{S_{t-1}^p}{N^p}\sum_{i=1}^{k}\beta_i^p(I_{t-iJ}^p - I_{t-(i-1)J}^p) + \delta\sum_q F(q,p)\frac{\sum_{i=1}^{k}\beta_i^q(I_{t-iJ}^q - I_{t-(i-1)J}^q)}{N^q}$$

Community spread

Travel spread



$$\beta^p = \begin{bmatrix}\beta_1^p & \dots & \beta_k^p & \delta^p\end{bmatrix}$$

$$\text{And, } \mathbf{X}_t^p = \begin{bmatrix} S_t(I_t^p - I_{t-J}^p) \\ \vdots \\ S_{t-(k-1)J}(I_{t-(k-1)J}^p - I_{t-kJ}^p) \\ \sum_q \frac{F(q,p)}{N^q}(I_t^q - I_{t-kJ}^q))]^T . \end{bmatrix}$$

$$\Delta I_t^p = \beta^p \mathbf{X}_t^p$$

Learning with weighted least square minimization

$$\sum_{t=1}^{T}(\alpha^{\frac{T-t}{2}}\Delta \hat{I}_t^p - \alpha^{\frac{T-t}{2}}\ \beta_p \mathbf{X}_t^p)^2$$

Decaying weights on past data

Results: Short-term Predictions (1)

- Using data by April 10th (not including travel)

	Method	RMSE (US)	MAPE (US)	RMSE (Global)	MAPE (Global)
Adaptive	SI-kJα (variable)	333.3	6.82%	462.6	13.64%
	SI-kJα (fixed)	342.05	6.58%	456.0	**11.22%**
	SI-kJα (ensemble)	**316.3**	**5.93%**	**355.9**	11.37%
Single curve fitting	Gen-SEIR	2106.4	14.31%	7471.2*	41.06%*

- Using data by March 21st including travel data

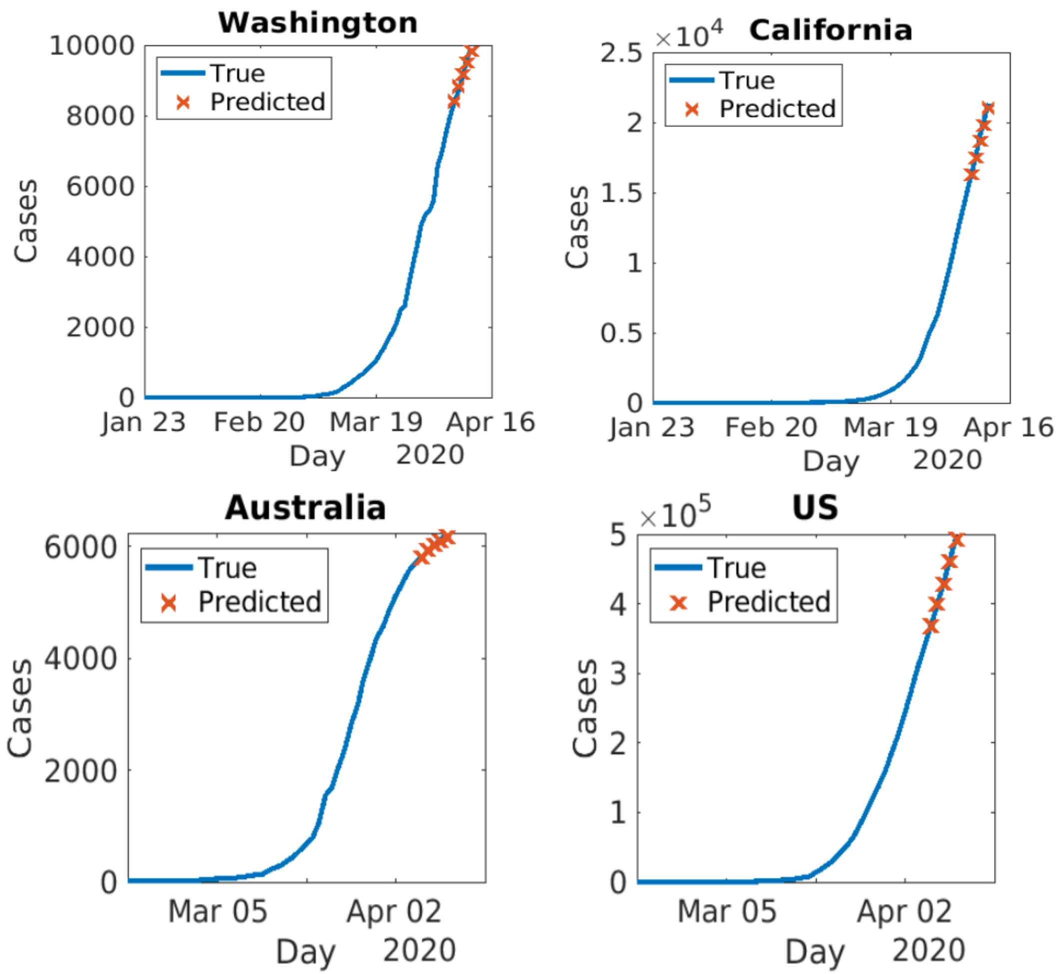
Travel data improved the models

Method	US RMSE	US MAPE	Global RMSE	Global MAPE
travel, variable	**147.3**	19.93%	248.4	21.353%
without travel, variable	166.7	**18.51%**	348.2	23.15%
travel, fixed	207.0	25.08%	**242.6**	**19.50%**
without travel, fixed	186.6	19.52%	286.8	21.42%

USC Viterbi School of Engineering

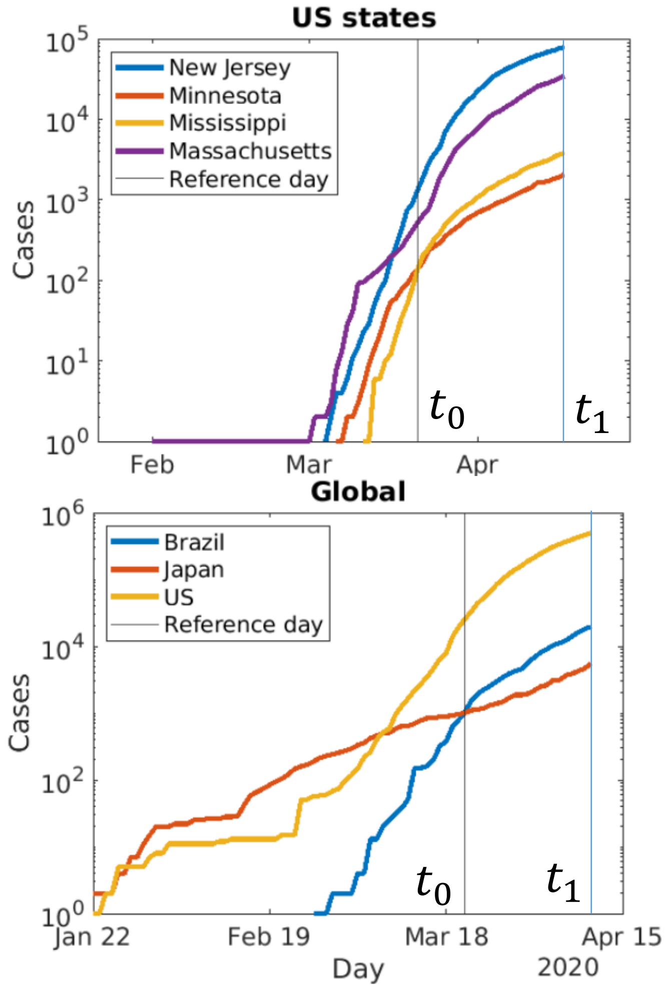
University of Southern California

Results: Short-term Predictions (2)



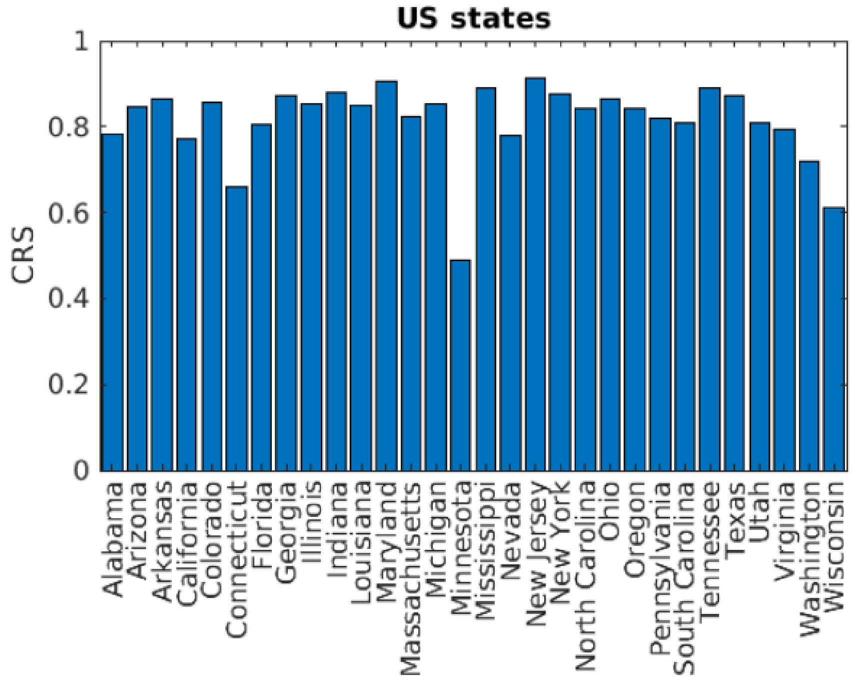
Measuring the Present, using the Past, through Predictions

- Compare
  - Reference day t_0 in the past– model parameters (M_0), forecast number of cases to the present ($I_0 \to^{M_0} I_1$)
  - Present day t_1 to measure – model parameters (M_1), actual confirmed cases on the present day ($\hat{I}_1$)

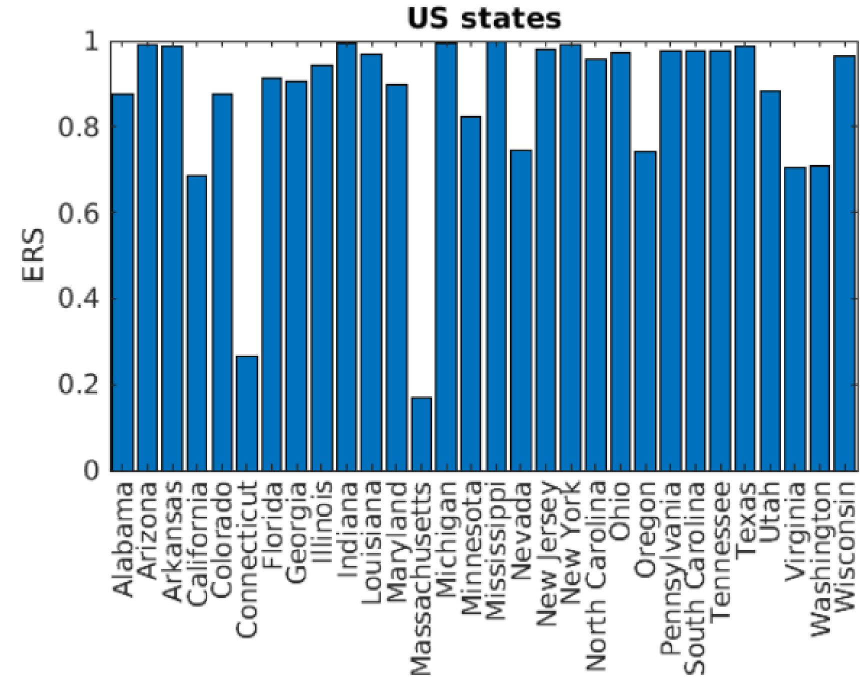


- We propose
  - **Contact Reduction Score (CRS):** A measure of reduction in transmission (M_0, M_1)
  - Depends only on model parameters
  - **Epidemic Reduction Score (ERS):** A measure of reduction in the number of cases
  - Depends on number of infections ($\hat{I}_1$, I_1)

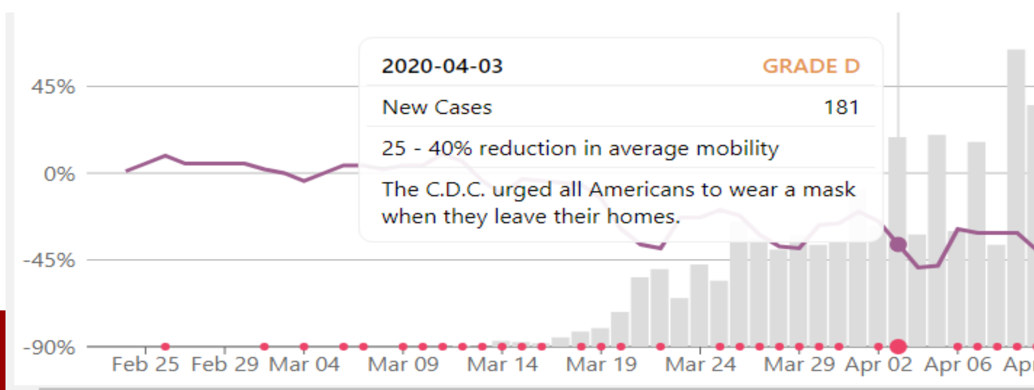
CRS and ERS for US States (March 21st-April 10th)



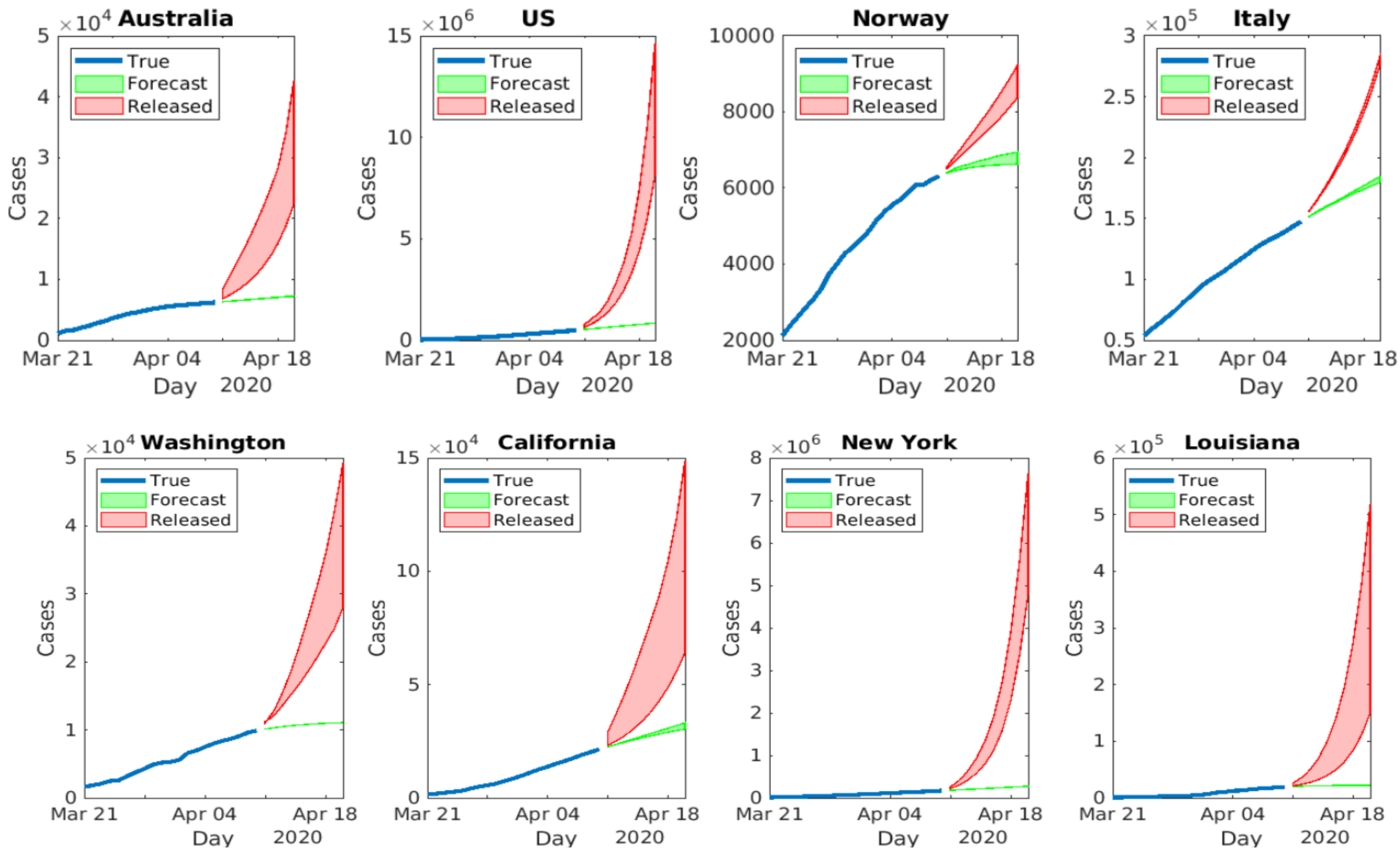
Best CRS: New Jersey, Worst CRS: Minnesota



Best ERS: Mississippi, Worst ERS: Massachusetts



Forecasts and "What-if"



Forecasting Web Interface

https://jaminche.github.io/COVID-19/

Conclusions

- Accurate short-term country and state-level forecasts
- Good hyperparameter selection is critical
- Models should evolve with data
- Ensemble approach likely to be the best approach
    - Consider several models instead of one
- Aggregate mobility reduction may not be the best way to rate the response

Next Steps

- County/city/neighborhood level predictions
- Hybrid hyperparameter/parameter learning scheme
  - Current approach: Each has its own or everyone uses the same hyperparameters
  - Clusters of regions share hyperparameters and even parameters: Consider similar regions when data for given region is not enough
- Incorporating Unreported Cases

Our Tools/Expertise for Other COVID-19 Researchers

- Resource allocation algorithms
  - Using the forecast to formulate and solve resource management problems [Bistra Dilkina, ...]
- Network diffusion/immunization
  - How to limit mobility so the epidemic is contained [Kristina Lerman, …]
- GraphSAINT (ICLR 2020): State-of-the-art Graph Embedding
  - Knowledge base for COVID-19 [Pedro Szekely, Jay Pujara]
  - Identifying candidate vaccines; effect on tissues [Barabasi Lab]

Acknowledgments

- **NSF RAPID:** ReCOVER: Accurate Predictions and Resource Allocation for COVID-19 Epidemic Response

- Initial Sprint
  - Frost Tianjian Xu (Sophomore, CS): Dataset preparation
  - Jamin Chen (Senior, CS): integrating our methods into a web-based visualization
  - Prathik Rao (Junior, CE) and Kangmin Tan (Junior, CS): Implementing and evaluating various ML training approaches

Data Science Lab led by Prof. Viktor K. Prasanna

Applied to: Social Good, Energy, System Design, Crime, …

Topics: Network Diffusion, Graph Analytics, Timeseries prediction, ML-driven Prefetchers, ML-driven Compilers, Smartgrids, Parallel Computing

Skills: Algorithms, Network Science, Data Mining, Math, Abstraction

We are looking for students!

Google Faculty Research Award

DARPA HIVE, SDH, PAPPA

NSF RAPID, SPX, EAGER, OAC

Intel NSF FoMR

Questions?

dslab.usc.edu

ajiteshs@usc.edu
prasanna@usc.edu