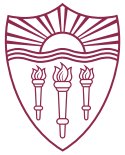


The EpiBench Platform to Propel AI/ML-based Epidemic Forecasting:

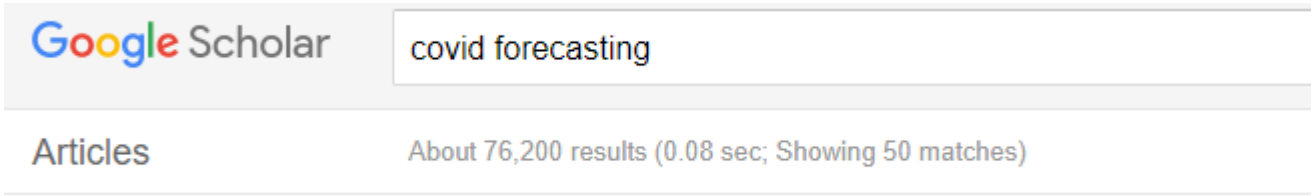
A Prototype Demonstration Reaching Human Expert-level Performance

Ajitesh Srivastava, Tianjian Xu, Viktor K. Prasanna

USC University of Southern California

Background

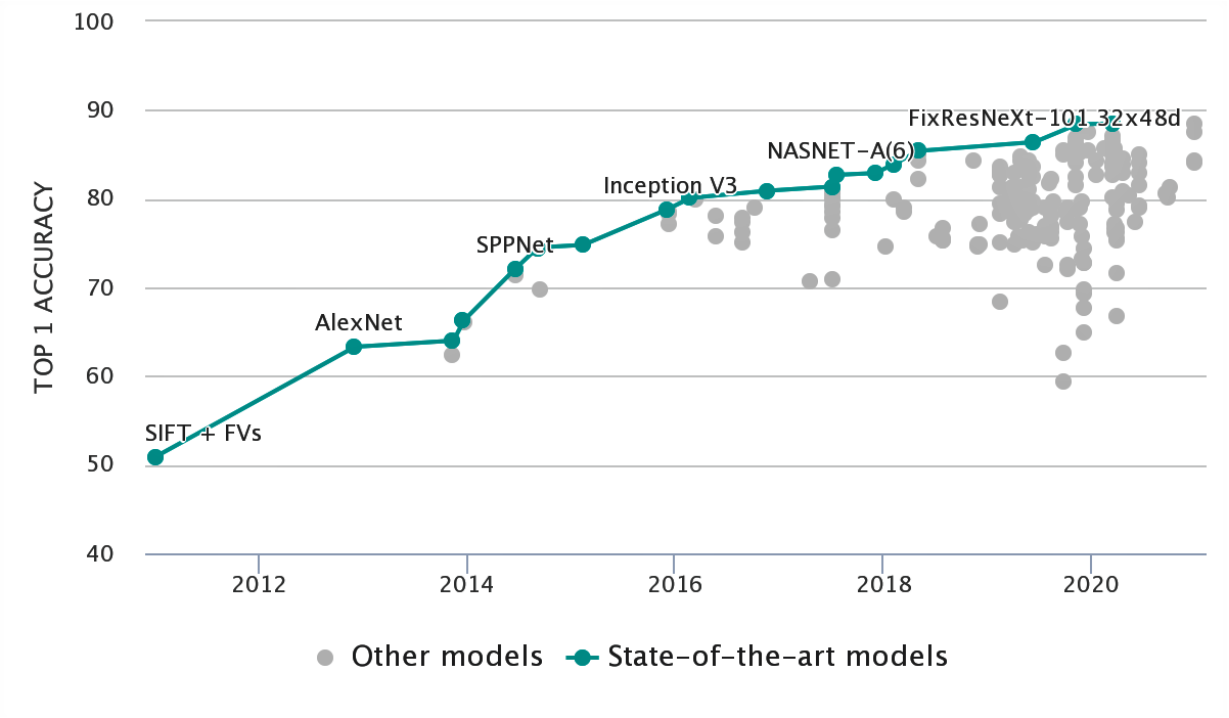
- “covid forecasting” on Google Scholar: ~76,000 results



- Claims at recent AI/ML conferences: AAAI, IEEE BigData
- AI/ML claims need proper evaluation
  - “state-of-the-art”
  - Reproducibility (without human intervention)

Benchmarks Help Push Boundaries

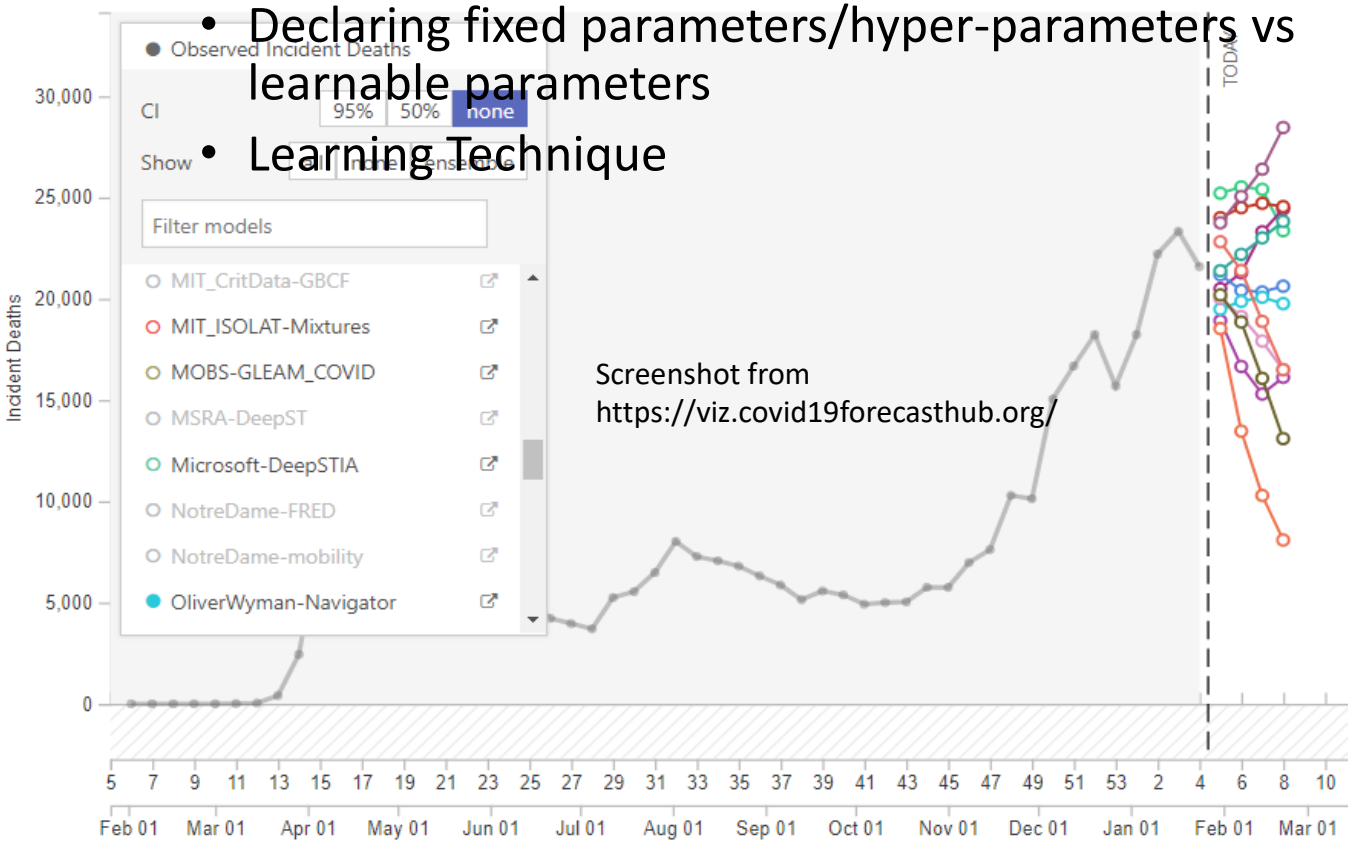
- In other domains
  - Image Classification: ImageNet/CIFAR-10/MNIST + MobileNet/VGG/…



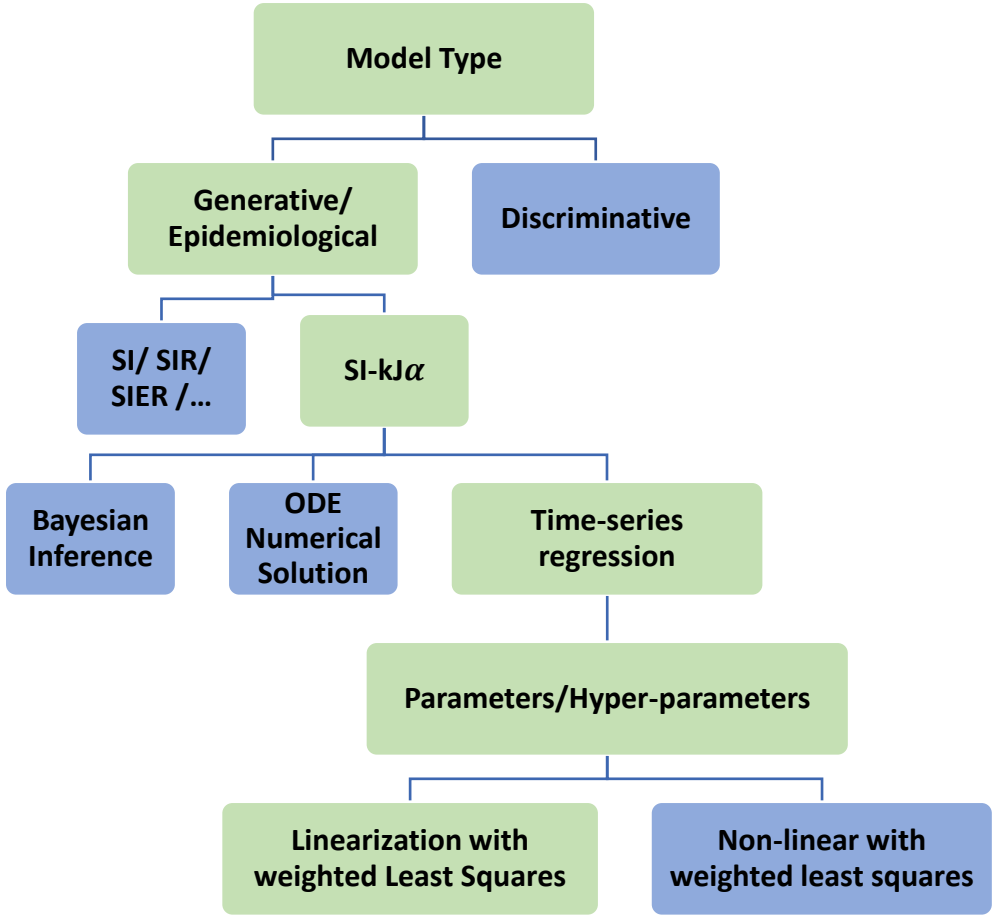
Background

"Our forecasting method outperforms SIR/SEIR"

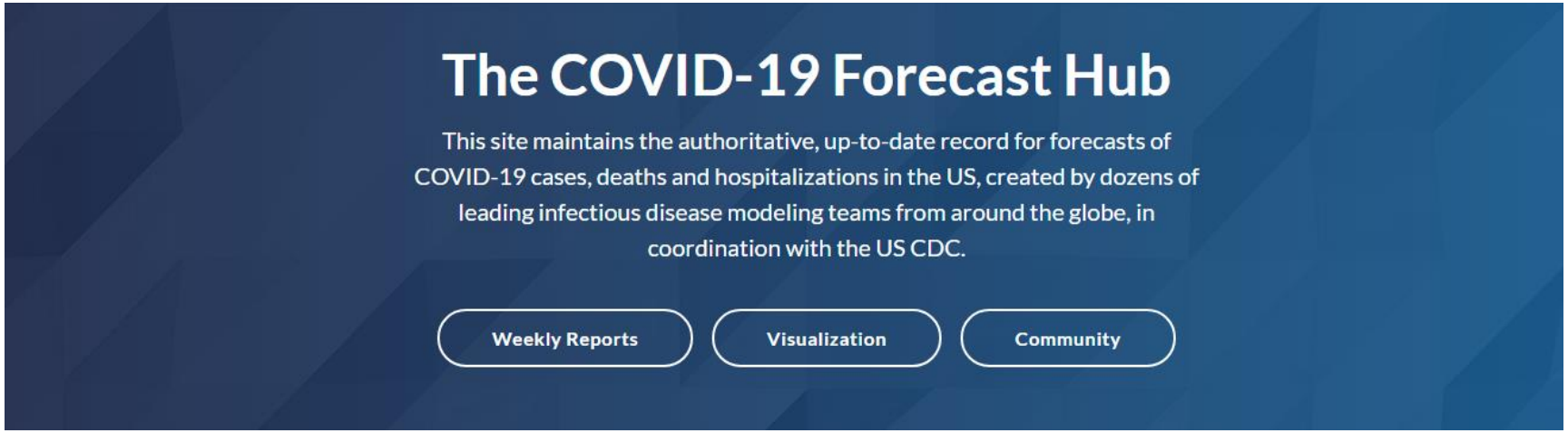
- **Model is not everything:** The path from data to forecasts involves more
  - Data pre-processing technique
  - Declaring fixed parameters/hyper-parameters vs learnable parameters
  - Learning Technique



Screenshot from https://viz.covid19forecasthub.org/

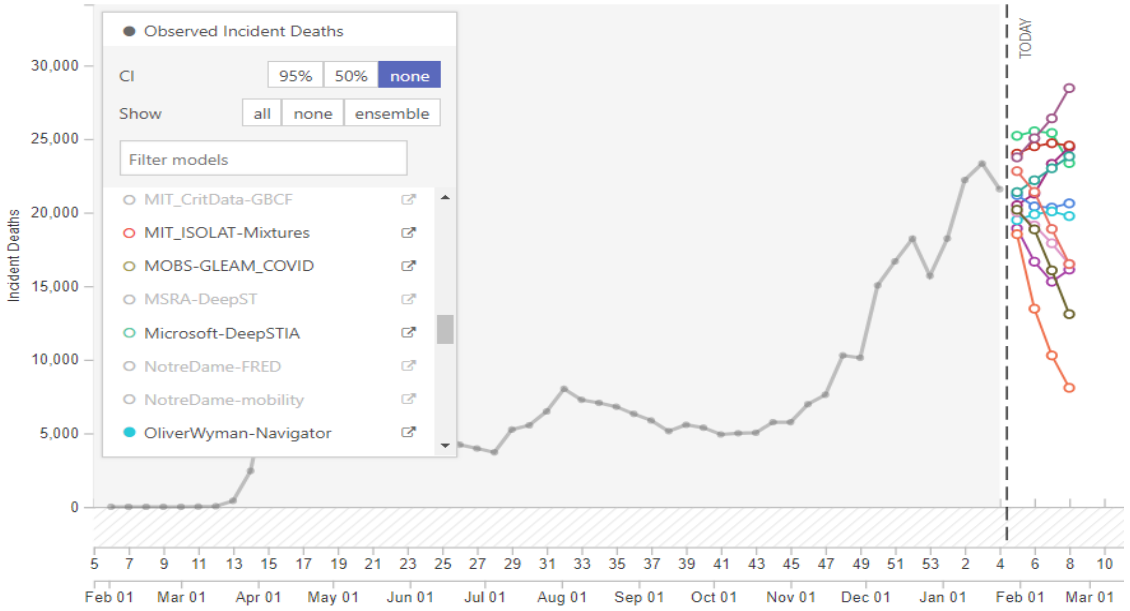


Existing Platforms: COVID-19 US Forecast Hub



Other platforms:

- COVID-19 Germany+Poland Forecast Hub (ECDC)
- CDC Epidemic Prediction Initiative

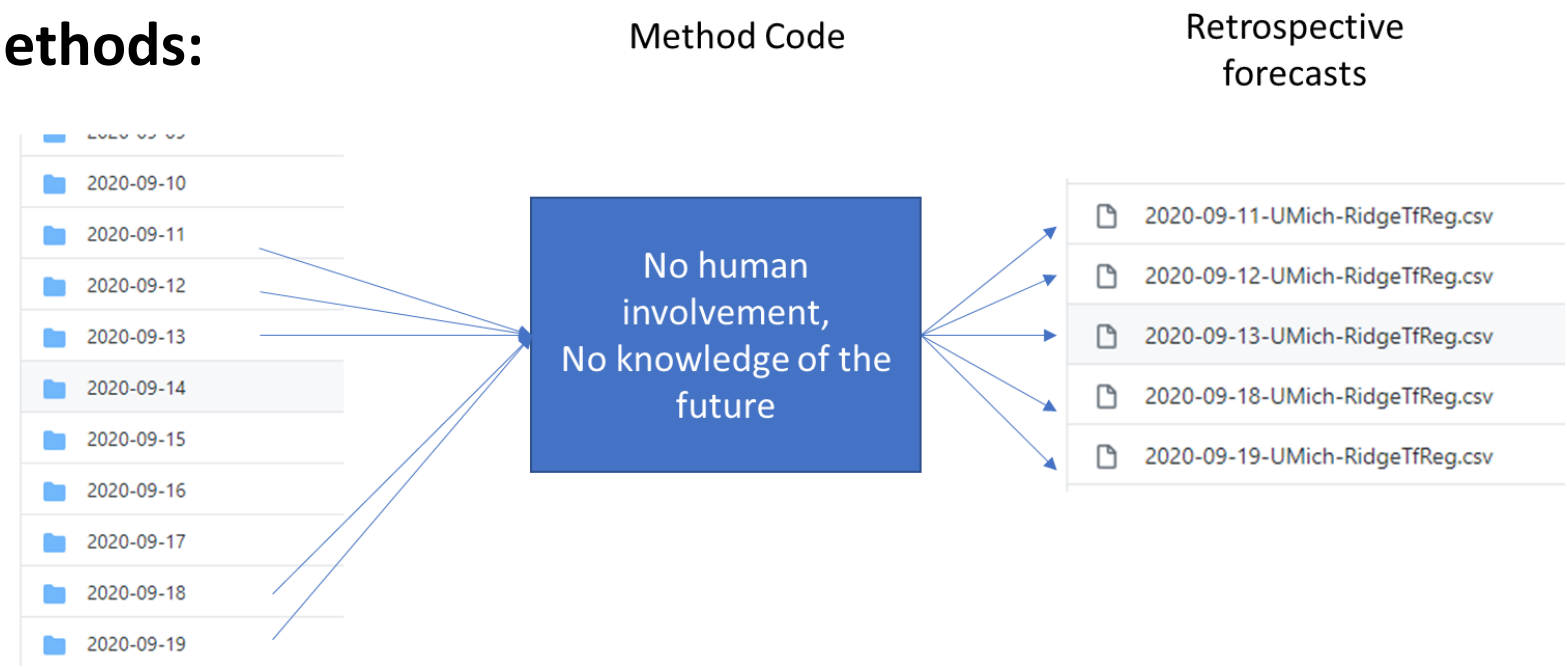


- ***Forecast-hub et. al.:*** *Provide the best guess for the future for an ongoing epidemic*
- ***Forecasting benchmark:*** *Provide performance reference for future ML methodologies*
  - MAE per region, mean MAE, …

We are Pushing for AI/ML vs “Human” Methods

- **Human Methods:** For the ongoing epidemic forecasting, periodically “experts” update forecasts based on intuition, changing data, …
  - Not a scalable approach to thousands of locations
  - Expertise stays with the expert

- **AI/ML methods:**



Long-term Goal

- To enable ML-driven research for epidemic forecasting by providing a collection of benchmarks (datasets + methods) that reflect the performance of various methodologies
  - For new methods to compare against
    - *Which method is state-of-the-art?*
    - *Can AI/ML with no human intervention outperform human-expert forecasts?*
  - To allow a more robust ensemble learning

This work is …

- Not claiming *"My method is the best epidemic forecasting method"*
  - But to enable such a claim
  - And to enable building better methods

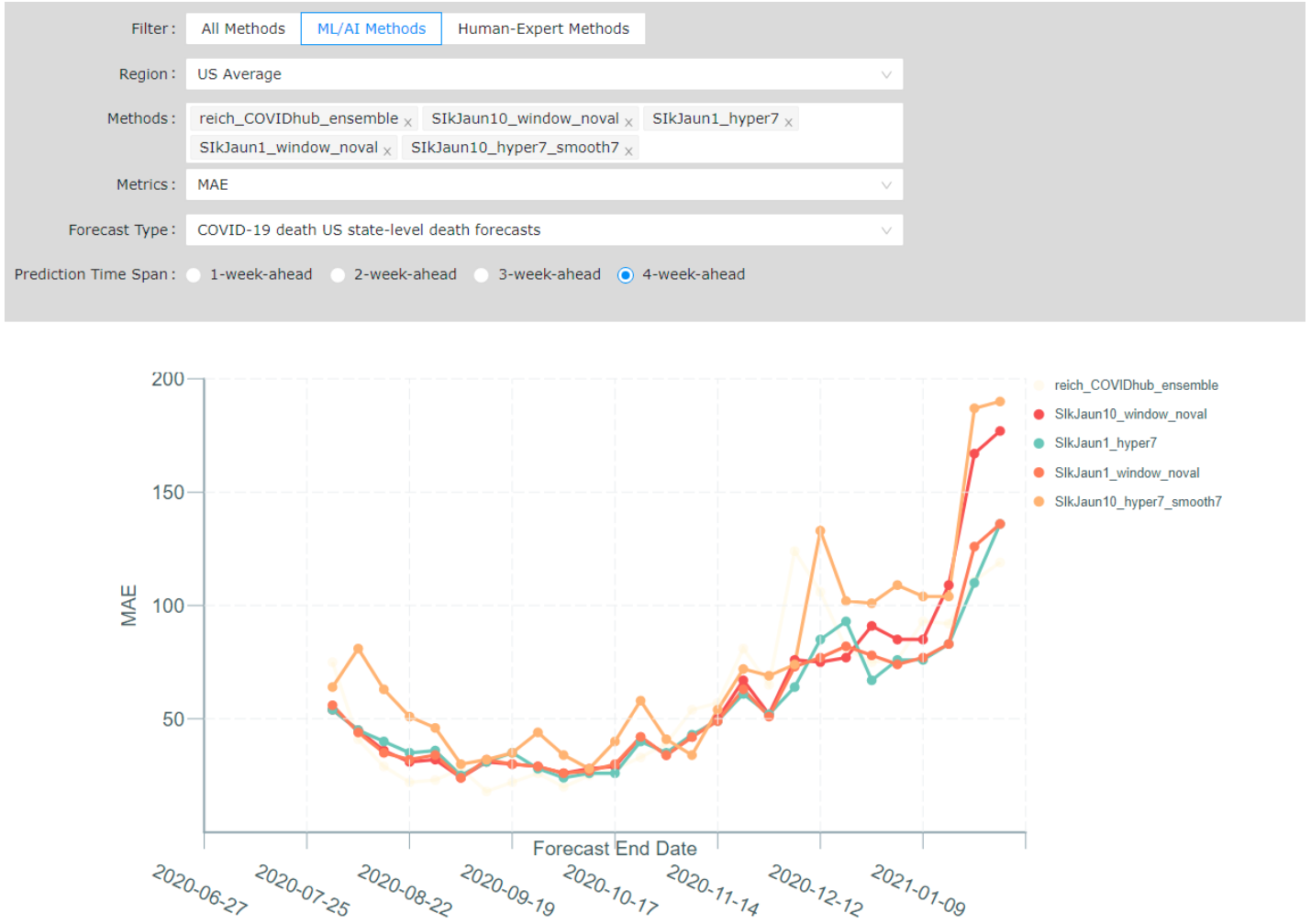
- Not claiming *"I have the best evaluations"*
  - Many forecasting teams are working on evaluating ongoing COVID-19 forecasts, including us in a different project
  - How to properly evaluate is still an open problem – We hope the platform will enable this discussion too

Contributions

- Introducing a prototype: **COVID-19 Forecasting Benchmark**
  - A platform for the task of COVID-19 case and death forecasts
  - Already accepting submissions

- Demonstrate that the platform can enable fully automated (no human intervention) ensemble reaching human-expert-level performance

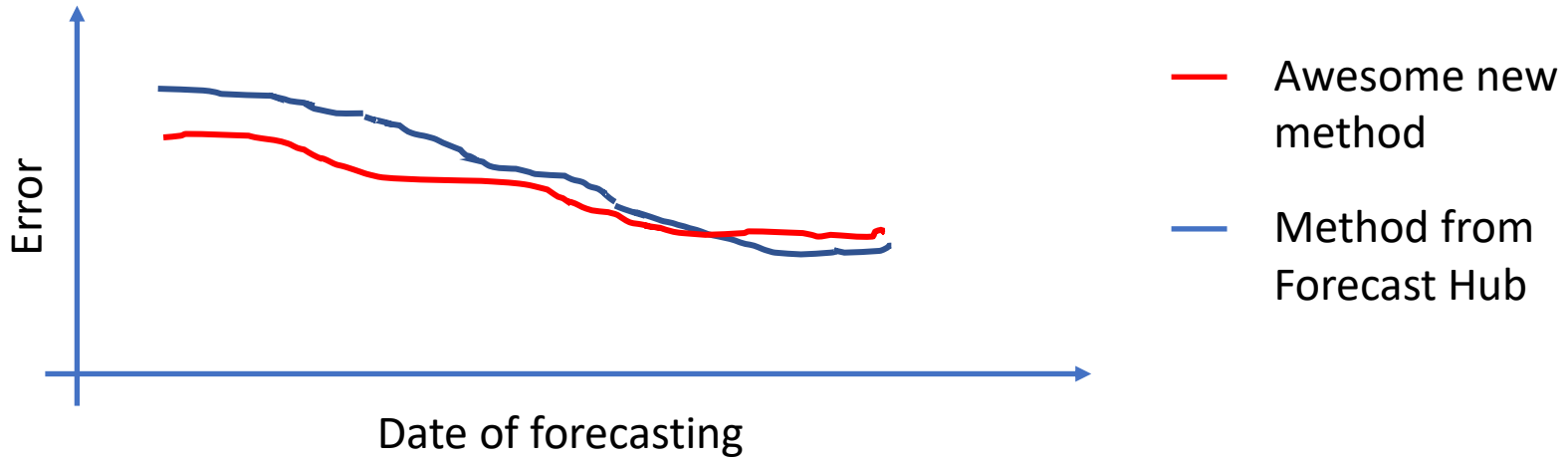
Comparison for New Methods



Our platform enables submissions of forecasts and comparisons of methods in an interactive way

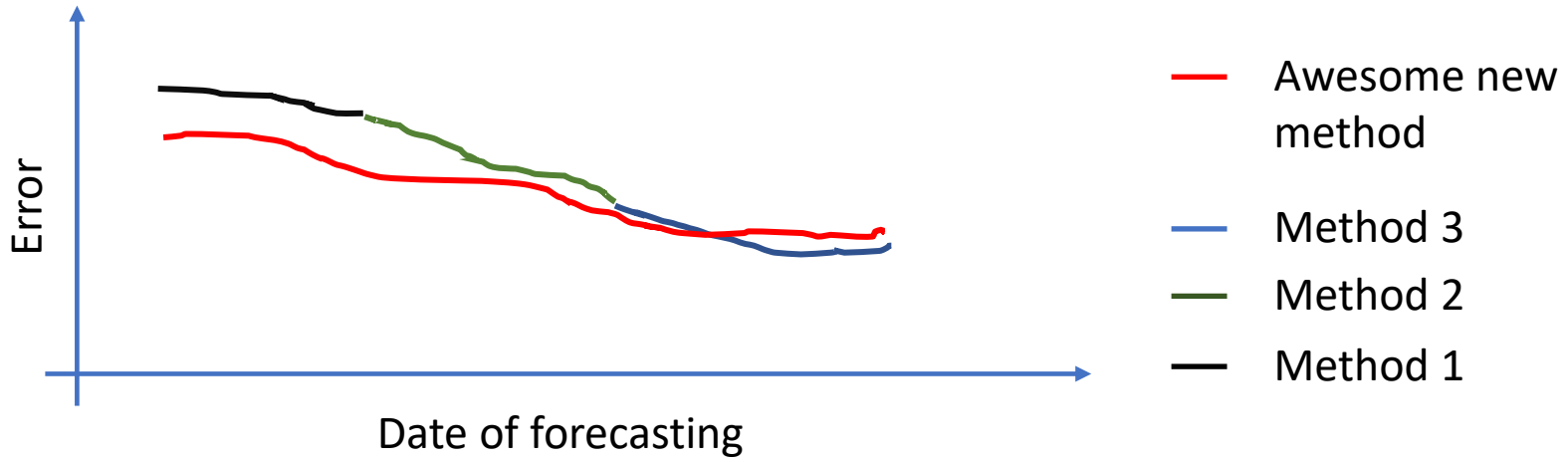
Comparison for New Methods

- Comparison against Forecast-hub?



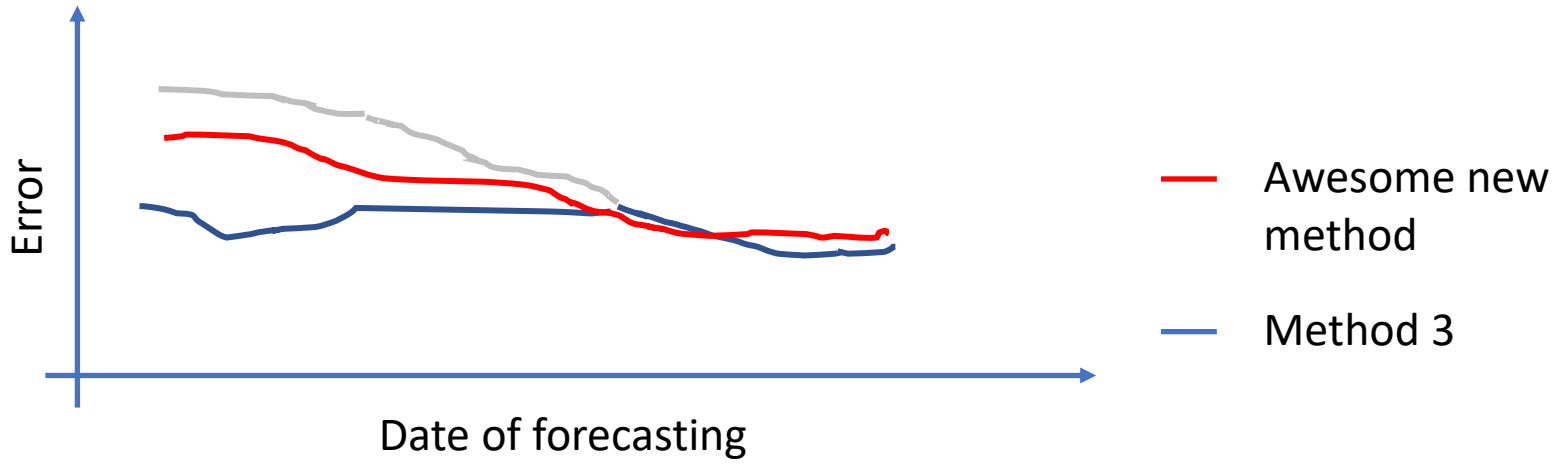
Comparison for New Methods

- Comparison against Forecast-hub?



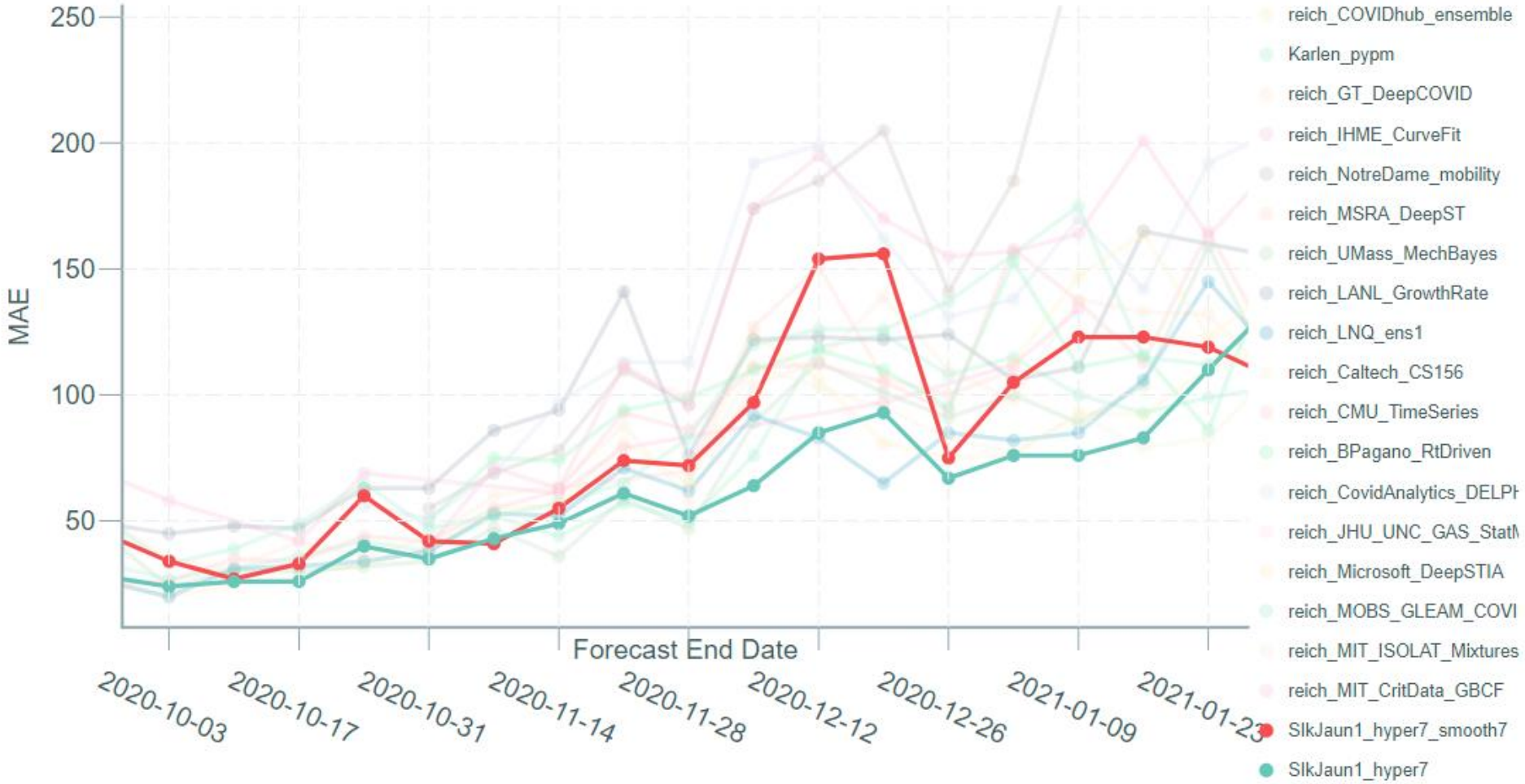
Comparison for New Methods

- Comparison against Forecast-hub?



“Which team is better” vs “Which method is better”

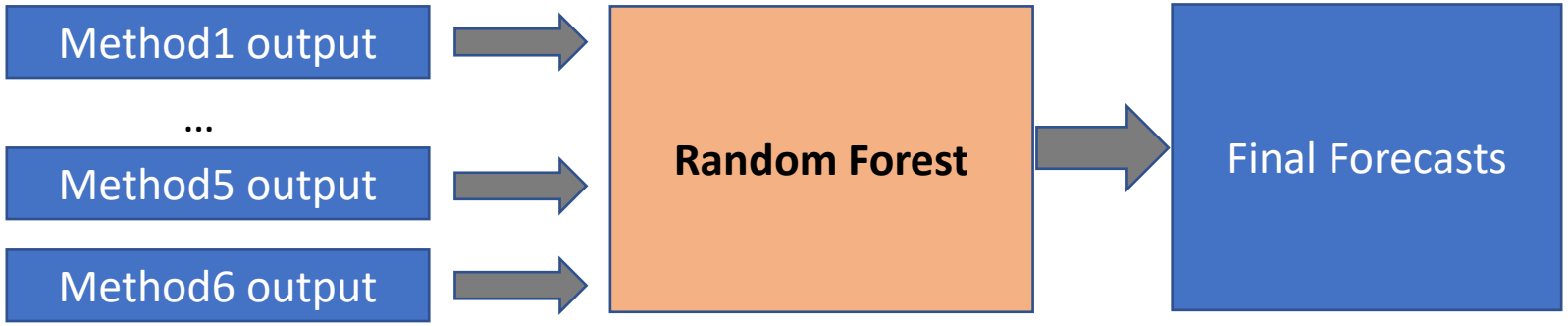
Tuning => Different method!



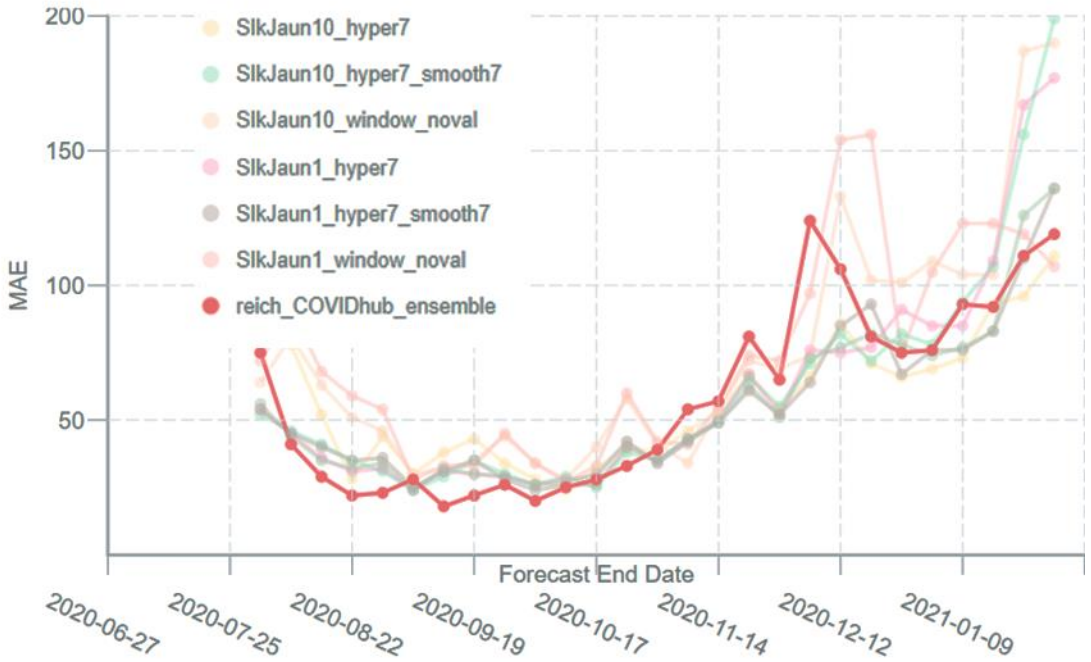
Ensemble Learning

- Current ensembles still rely on simple averaging/median of individual method (constituent) forecasts
  - Among evolving methodologies, a smarter rule is difficult to learn
- With **consistent** methodologies, we can learn hidden rules on how to combine the results

- Example Approach:



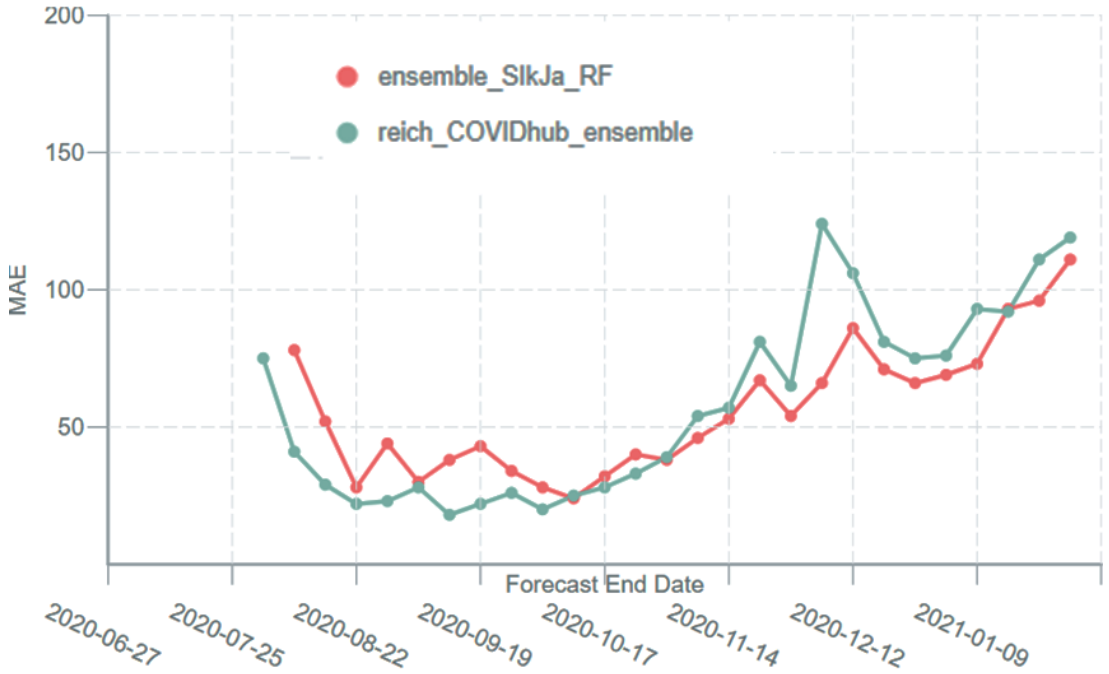
Ensemble Results: 4 week ahead incident death forecasts



- Constituent methods: Variations of "SIkJalpha" :
  - Various settings for pre-processing (e.g., smoothing factor)
  - With or without "validation set"

- **Observation:** Different constituent methods better or worse at different points in time compared to "COVIDhub Ensemble"

Ensemble Results: 4 week ahead incident death forecasts



- **Observation:** Random Forest Ensemble close to COVIDhub-ensemble, in many cases better!
- Recall:
  - **COVIDhub Ensemble:** Combination of forecasts from > 30 expert teams each tuning their methods over time
  - **This Ensemble:** Combination of fast forecasting methods, all *without human intervention*

Submissions

- Help us create this resource for AI/ML community!
- Details at: https://scc-usc.github.io/covid19-forecast-bench
- Thanks to those who have already submitted and/or provided details of methodology!