Carnegie Mellon University
Statistics & Data Science

OPTUM

May 12, 2021

Optum | Project Presentation

Carnegie Mellon University Statistics & Data Science

Amy Lopata, Carol Yin, Steve Wong

Agenda

1 Research Question

2 Literature Review

3 Interactive Graphics App Demonstration

Optum

Healthcare service and innovation company on a mission to help people live healthier lives and to help make the health system work better for everyone.

A part of UnitedHealth Group, formed by merging pharmacy and care delivery services into the single Optum brand, comprising three main businesses: OptumHealth, OptumInsight and OptumRx.

Research Question

With the arrival of COVID-19, forecasting cases are necessary for planning, like moving hospital supplies and tracking the effectiveness of interventions.

Using statistics methodology, which forecasting model or combination of models published by other COVID-19 research groups best predicts cases and hospitalization over time?

Project Overview

Literature Review of 26 COVID-19 forecasting models

Interactive Graphics Web App

- Evaluate and visualize models’ performance by state and forecast horizon
- Currently includes forecasts from Dec. 2020 - Feb. 2021

Literature Review

Objectives

Investigate features of COVID-19 forecasting models published by the CDC

- 26 COVID-19 research groups of interest (identified by Optum)
- Features of interest include:
  - Group affiliation
  - Geographic resolution (county, state, national)
  - Forecast types (new cases, hospitalizations, deaths)
  - Methods
  - Modeling assumptions
  - Performance measures
  - Data sources
- Each model provides cumulative forecasts for 1, 2, 3, and 4 weeks ahead

Summary: Research Groups and Forecast Types

- Group affiliations
    - **University affiliated:** 16 groups
    - **Industry affiliated:** 4 groups (IEM, LANL, Microsoft, SignatureScience)
    - **Non-affiliated individuals:** 6 groups (BPagano, Karlen, LNQ, OneQuietNight, QJHong, ESG)
- Forecast Types
    - **Cases, hospitalizations, and deaths:** 8 models
    - **Cases and deaths only:** 13 models
    - **Cases only:** 4 models (IEM, OneQuietNight, UCF, UVA)
    - **Deaths only:** 1 model (BPagano)

Summary: Methods

- **Compartmental Models**
  - SEIR model (or modified SEIR): 8 groups
  - SIR model: 2 groups
- **Statistical/Machine Learning Models**
  - Machine Learning/Deep Learning Model: 5 groups
  - Time Series Related Model: 2 groups
  - Others (e.g., bayesian model, ridge regression model, statistical random walk model, model using gaussian distributions)
- **Ensemble Models**
  - 1 group uses SEIR + Neural Network
  - 1 group uses SEIR + Auto-aggressive Model + Machine Learning

Summary: Data Sources

- 13 groups use the Johns Hopkins Center for Systems Science and Engineering (JHU-CSSE) data
- 5 groups use the New York Times data
- 6 groups use the Covid Tracking Project Data
- 8 groups have ambiguous sources of data or did not release which data source they used
- 12 groups only used one data source in their model

Performance Measures

OneQuietNight - Normalized MAE (with comparison of other models)

Table 1. Forecasting accuracy of forecasts between 2020-09-14 and 2020-11-02. We computed the mean absolute error using the daily reports containing the cases data from the JHU CSSE group as the gold standard reference for the cases in the US. We normalized all the numbers by the COVIDhub-baseline number. The normalized mean absolute error numbers for each of the forecast horizons are shown below (lower is better).

	country				county				state			
target	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
OneQuietNight	0.72	0.93	0.97	0.91	0.95	0.95	0.93	0.93	0.86	0.83	0.84	0.88
COVIDhub-baseline	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CEID-Walk	1.00	1.00	1.00	0.99	1.01	1.01	1.01	1.00	1.00	1.01	1.01	1.01
CMU-TimeSeries	nan	nan	nan	nan	6.24	6.51	7.19	7.73	nan	nan	nan	nan
CU-nochange	0.99	0.70	0.66	0.62	1.19	1.23	1.32	1.45	1.09	1.01	1.02	1.04
CU-scenario_high	0.98	0.62	0.56	0.61	1.19	1.24	1.33	1.41	1.09	0.98	0.96	0.99
CU-scenario_low	1.03	0.83	0.84	0.87	1.19	1.24	1.31	1.37	1.11	1.08	1.12	1.19
CU-scenario_mid	0.98	0.66	0.75	0.75	1.19	1.23	1.25	1.23	1.09	0.99	0.99	0.98
CU-select	0.98	0.66	0.75	0.75	1.19	1.23	1.25	1.23	1.09	0.99	0.99	0.98
Columbia_UNC-SurvCon	0.61	0.53	0.63	0.87	nan	nan	nan	nan	nan	nan	nan	nan
Covid19Sim-Simulator	1.21	1.09	1.03	0.99	nan	nan	nan	nan	1.24	1.16	1.13	1.12
CovidAnalytics-DELPHI	2.91	1.96	1.62	1.46	nan	nan	nan	nan	2.68	1.97	1.72	1.64
DDS-NBDS	0.65	0.56	0.64	0.80	nan	nan	nan	nan	1.08	0.87	0.92	1.14
Geneva-DetGrowth	0.82	nan	nan	nan	nan	nan	nan	nan	0.89	nan	nan	nan

ISU - RMSPE (with comparison of two models discussed in the paper)

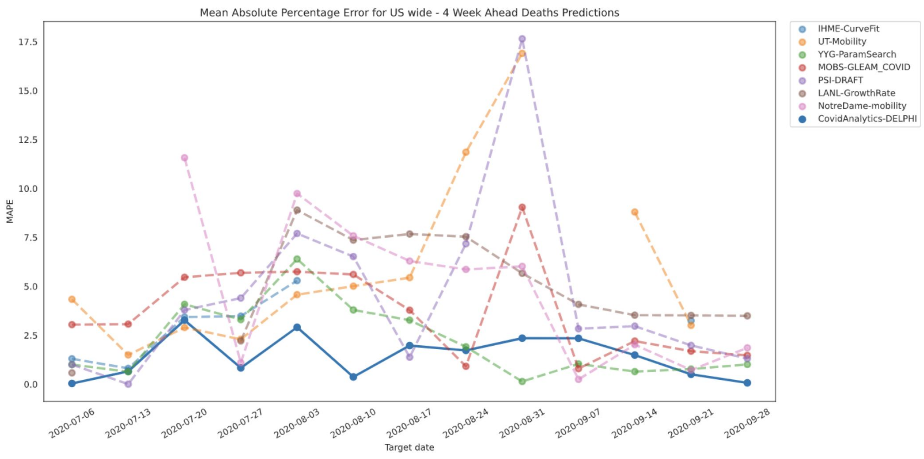
Table 6.2: The average of root mean squared prediction errors (RMSPE) of the infection or death count, where D_h is for the h-day ahead prediction, $h = 1, \ldots, 28$.

Infection Model										
Model	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
STEM	3.28	5.01	6.76	8.36	9.93	11.55	13.58	15.32	17.15	18.94
Linear	7.11	9.53	11.99	14.57	17.76	21.10	24.88	29.24	34.08	39.58
Model	D_{11}	D_{12}	D_{13}	D_{14}	D_{15}	D_{16}	D_{17}	D_{18}	D_{19}	D_{20}
STEM	20.91	22.83	24.78	26.68	28.76	30.86	32.93	35.20	37.34	39.54
Linear	45.71	52.23	58.98	66.09	73.27	80.55	88.03	95.72	103.75	112.25
Model	D_{21}	D_{22}	D_{23}	D_{24}	D_{25}	D_{26}	D_{27}	D_{28}		
STEM	41.77	43.94	46.17	48.47	52.68	55.42	58.27	61.22		
Linear	121.27	130.83	140.92	151.77	166.06	181.3	197.78	215.55		
Death Model										
Model	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
STEM	1.17	1.67	2.05	2.40	2.7	2.96	3.21	3.45	3.68	3.90
Linear	2.43	3.16	3.91	4.67	5.47	6.31	7.20	8.13	9.10	10.08
Model	D_{11}	D_{12}	D_{13}	D_{14}	D_{15}	D_{16}	D_{17}	D_{18}	D_{19}	D_{20}
STEM	4.08	4.29	4.49	4.69	4.87	5.04	5.18	5.33	5.47	5.60
Linear	11.06	12.11	13.21	14.32	15.46	16.66	17.84	19.11	20.39	21.69
Model	D_{21}	D_{22}	D_{23}	D_{24}	D_{25}	D_{26}	D_{27}	D_{28}		
STEM	5.78	5.93	6.10	6.26	6.42	6.59	6.76	6.95		
Linear	23.02	24.38	25.81	27.26	28.78	30.3	31.89	33.53		

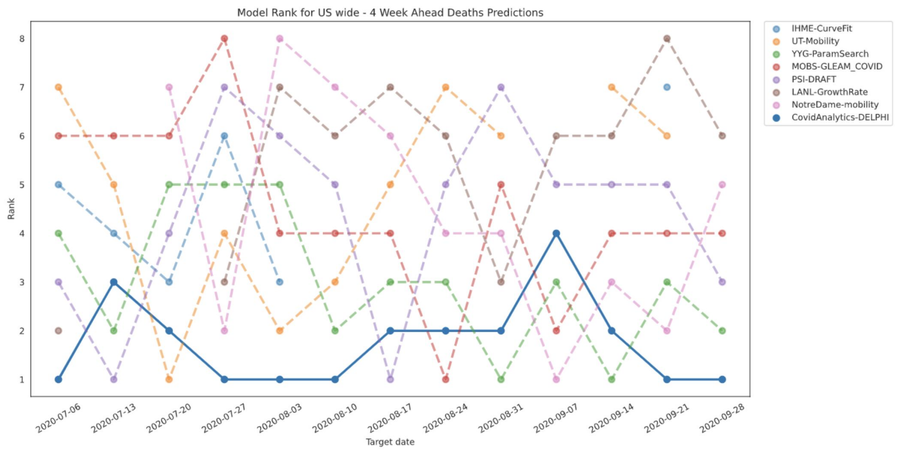
Performance Measures Continued

MIT ORC - Mean Absolute Percentage Error

- Comparison of other models
- Ranking graph



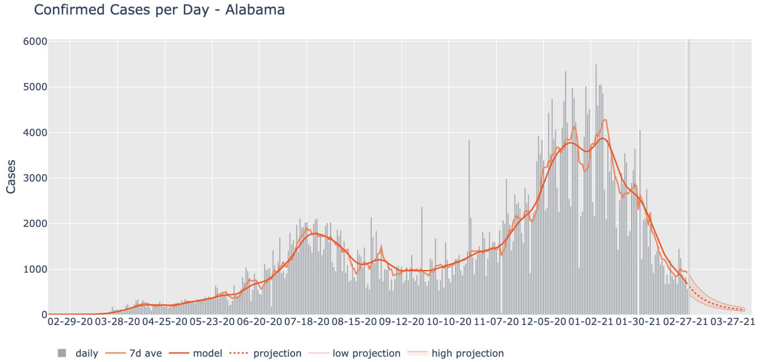
MAPE with comparison of other models

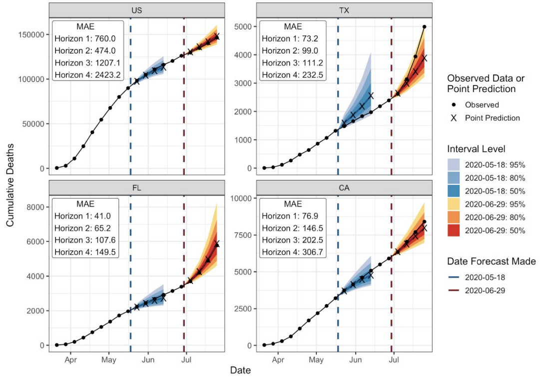


Ranking

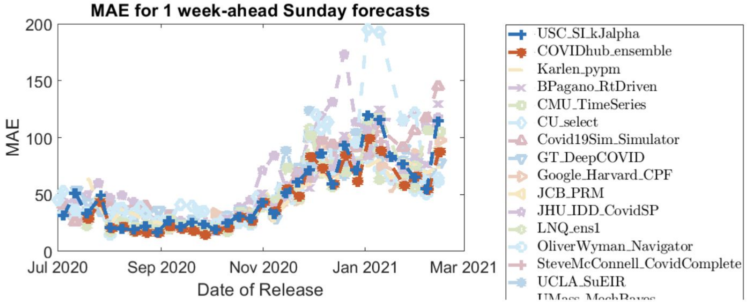
Performance Measures Illustration

Daily Accuracy (with Point Estimate)
+ Coverage Rate (with Confidence Intervals)





MAE by USC



RMSPE by ISU

Table 6.2: The average of root mean squared prediction errors (RMSPE) of the infection or death count, where D_h is for the h-day ahead prediction, $h = 1, \ldots, 28$.

Infection Model										
Model	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
STEM	3.28	5.01	6.76	8.36	9.93	11.55	13.58	15.32	17.15	18.94
Linear	7.11	9.53	11.99	14.57	17.76	21.10	24.88	29.24	34.08	39.58
Model	D_{11}	D_{12}	D_{13}	D_{14}	D_{15}	D_{16}	D_{17}	D_{18}	D_{19}	D_{20}
STEM	20.91	22.83	24.78	26.68	28.76	30.86	32.93	35.20	37.34	39.54
Linear	45.71	52.23	58.98	66.09	73.27	80.55	88.03	95.72	103.75	112.25
Model	D_{21}	D_{22}	D_{23}	D_{24}	D_{25}	D_{26}	D_{27}	D_{28}		
STEM	41.77	43.94	46.17	48.47	52.68	55.42	58.27	61.22		
Linear	121.27	130.83	140.92	151.77	166.06	181.3	197.78	215.55		

Information Spreadsheet:
https://docs.google.com/spreadsheets/d/1IWj3ecKVmLZNNWqh3uYqhKuUPbwW2TBvCgOcwc1ZEV4/edit#gid=0

Challenges

- Information is written for the public which can make it difficult to find documentation of technical details (e.g., performance measure, methods)
- Data sources are not readily available on the websites

Interactive Graphics App

https://qiuyiyin-shiny.shinyapps.io/Forecast-Performance/

Objectives

Compare model performance to address questions such as:

- Which models perform best in each state?
- Which models perform best in short-term and long-term forecasting?
- Do some methods (statistical/machine learning, compartmental, ensemble) perform better than others in certain conditions? Do some methods over/underestimate reported cases more often than others?
- How does model performance change over time, particularly before/after surges?

Model Performance by Error Type, State, Forecast Date

Forecast Performance

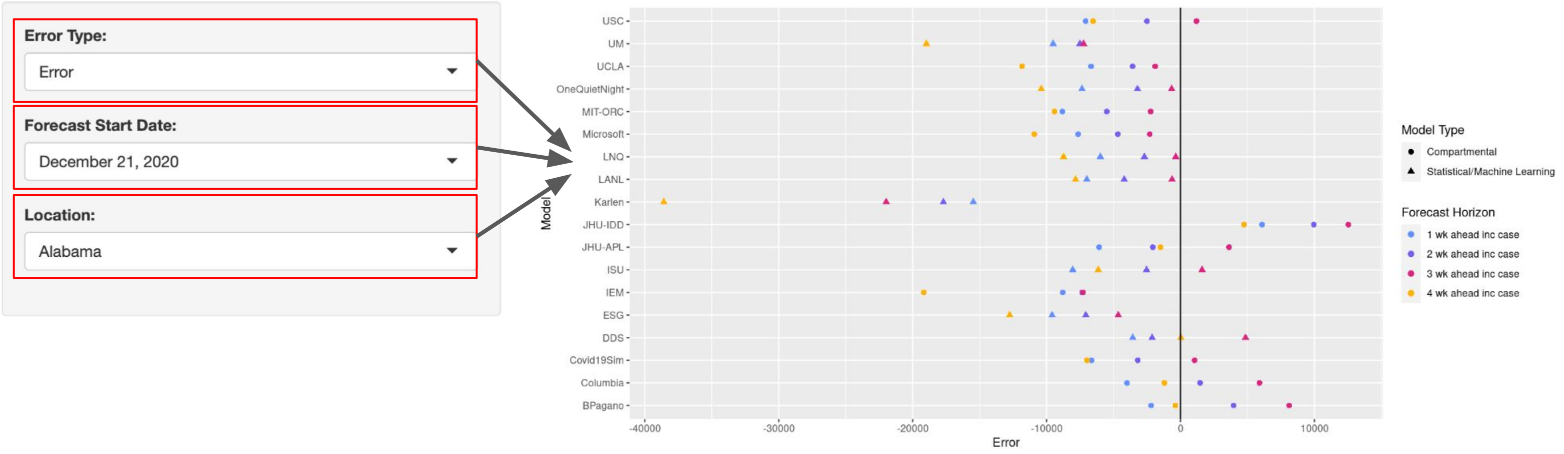
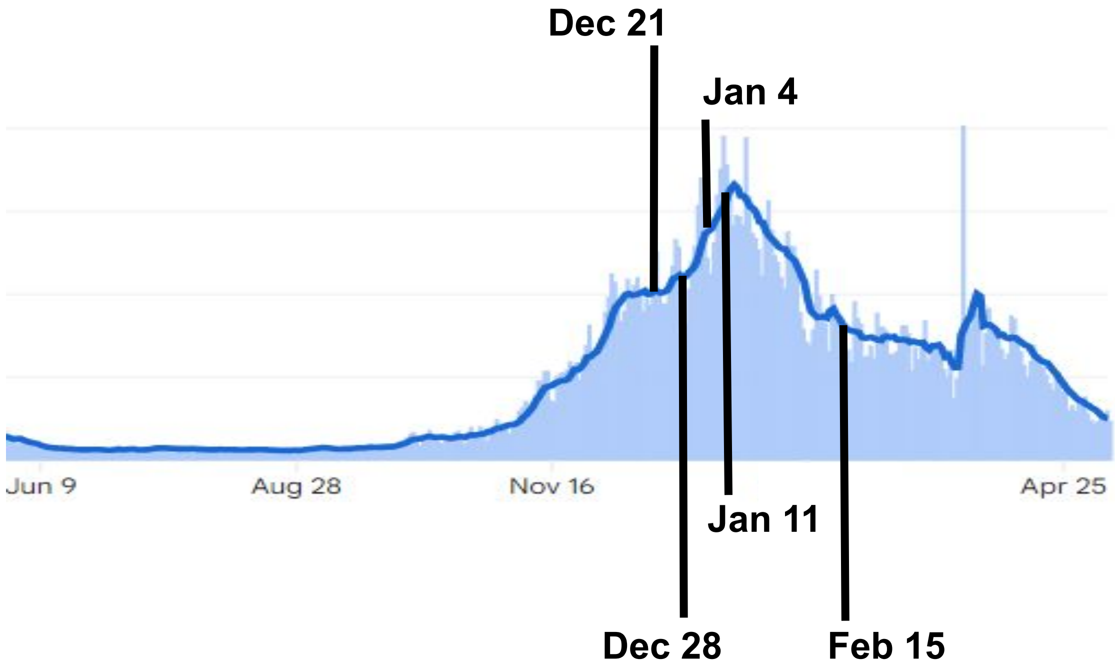


Table: Best Models by Forecast Horizon

1 wk ahead inc case	2 wk ahead inc case	3 wk ahead inc case	4 wk ahead inc case
BPagano	Columbia	LNQ	DDS
DDS	JHU-APL	LANL	BPagano
Columbia	DDS	OneQuietNight	Columbia

https://qiuyiyin-shiny.shinyapps.io/Forecast-Performance/

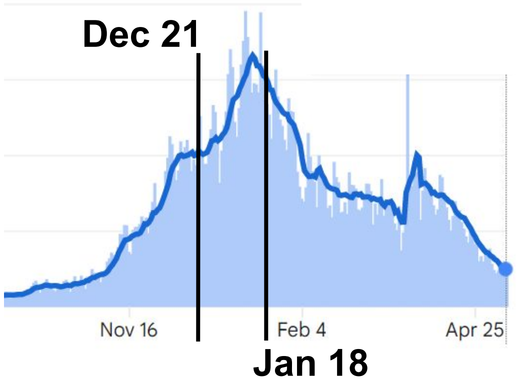
Example: New York Reported Cases

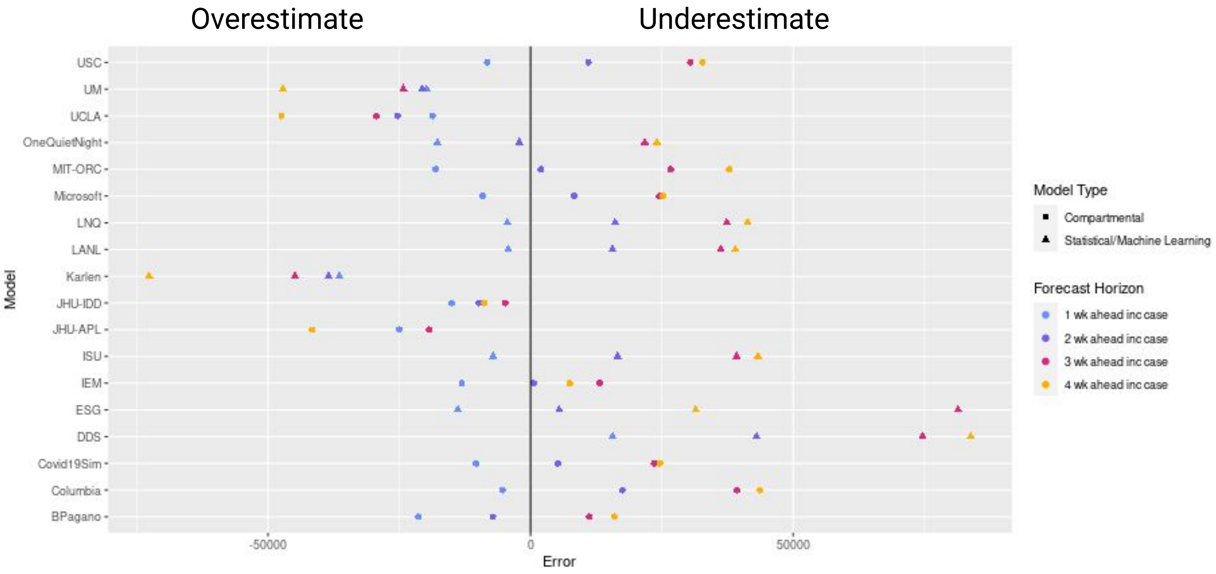


Source: The New York Times

Example: New York December 21st

Reported Cases

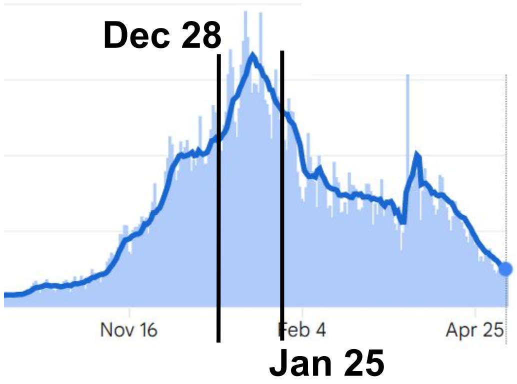


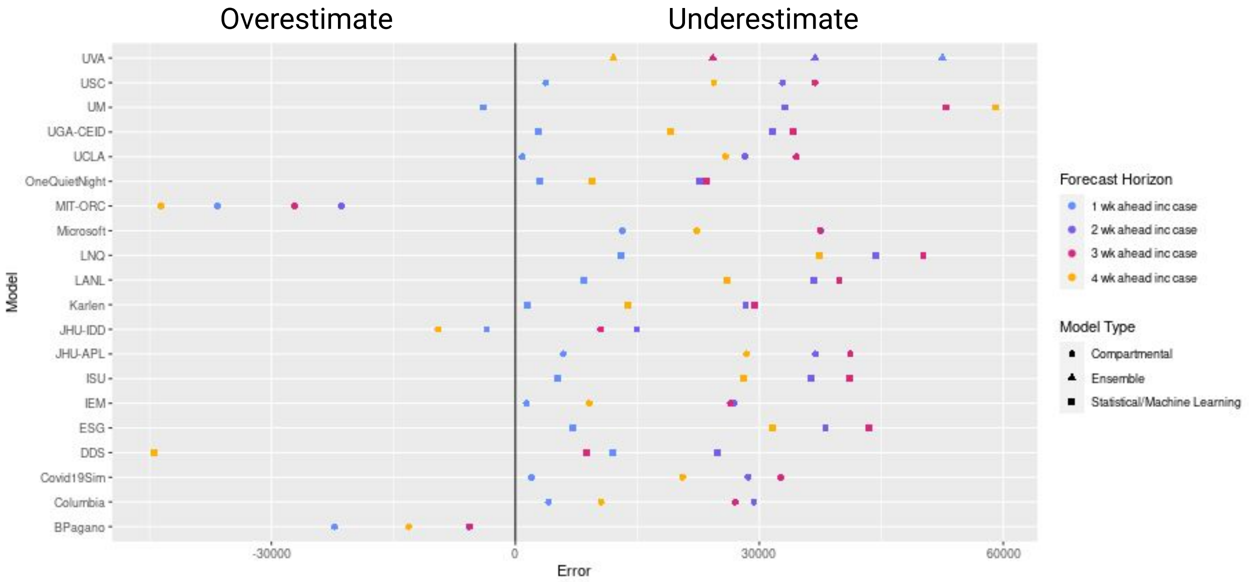


Most models **underestimated** the number of new cases for 3 and 4 weeks ahead.

Example: New York December 28th

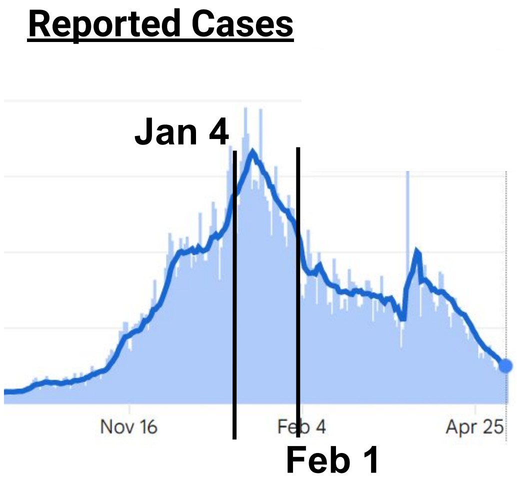
Reported Cases

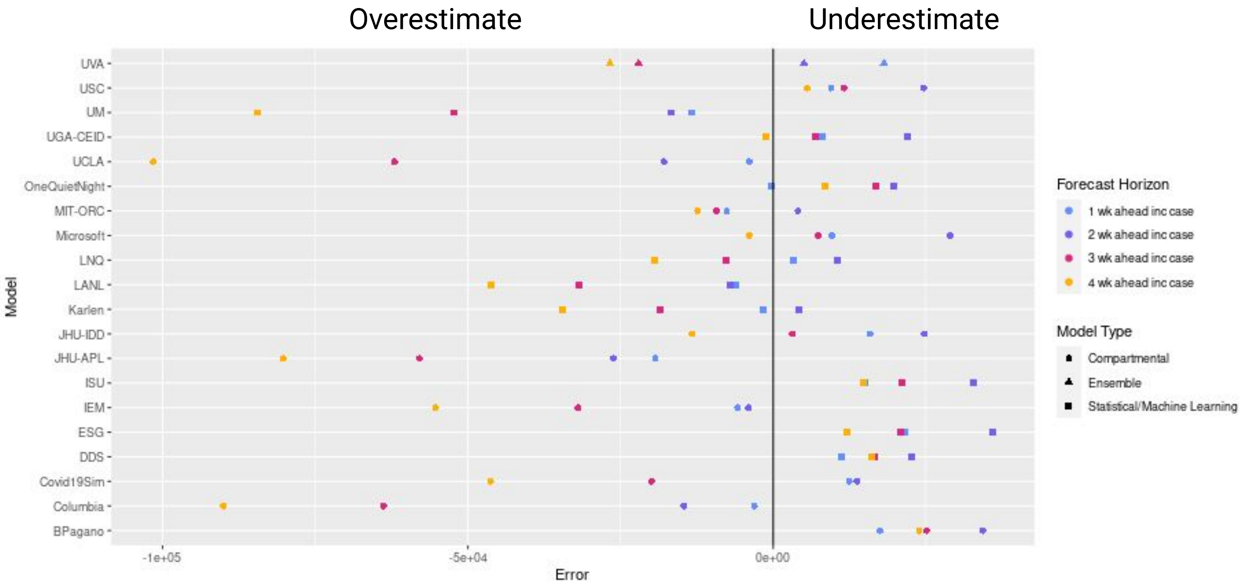




Most models **underestimated** the number of new cases during this surge.

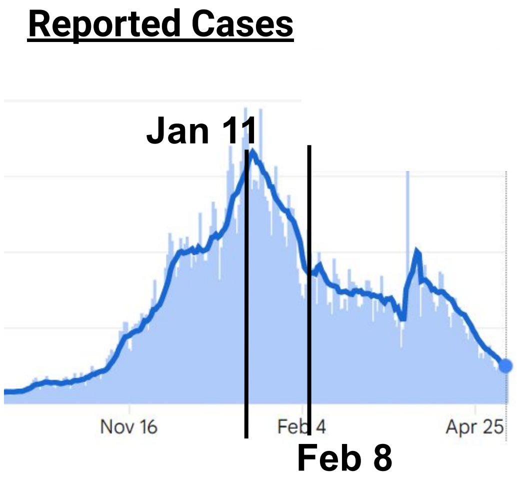
Example: New York January 4th

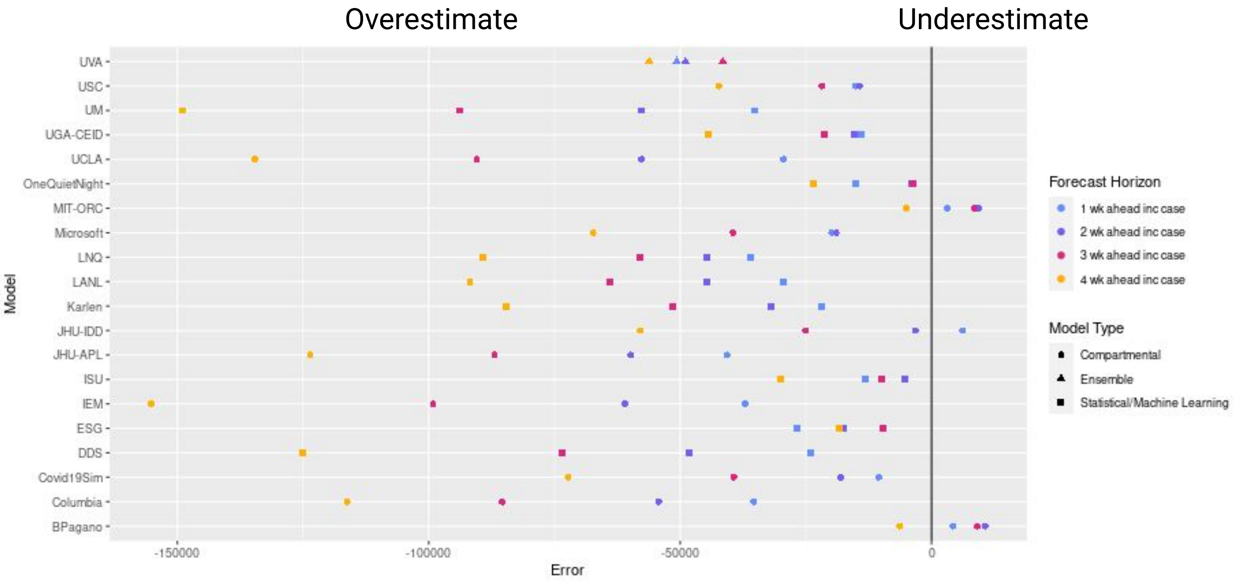




Many models **overestimated** the number of new cases for 3 and 4 weeks ahead.

Example: New York January 11th

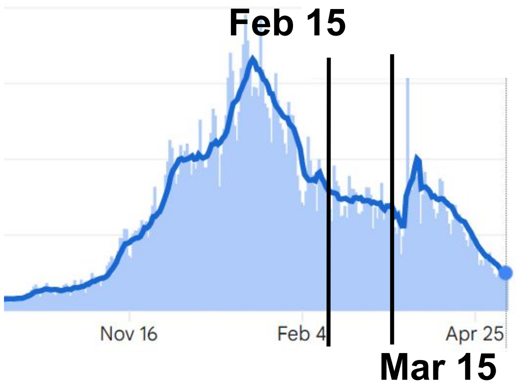


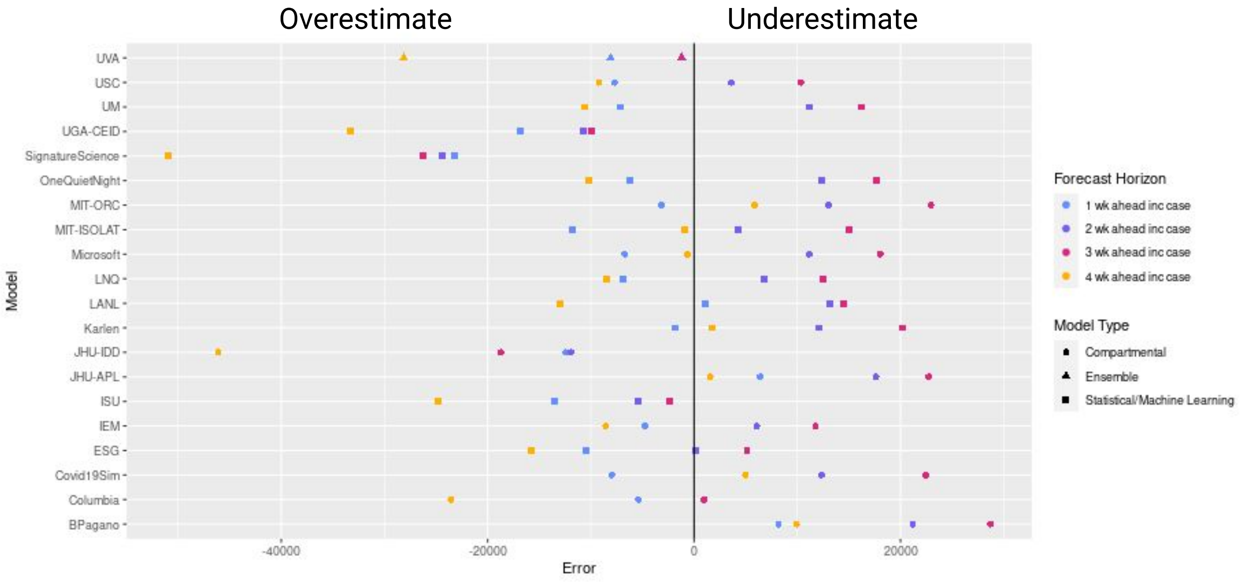


Most models **overestimated** the number of new cases after this surge.

Example: New York February 15th

Reported Cases





During a plateau following the surge, some models overestimated and some models underestimated the number of new cases.

Summary: Top Models for February 15th Forecasts

- Ranking methodology
  - For each state at each forecast horizon, assigned 3 points to the 1st place model (model with lowest absolute error), 2 points to 2nd place, and 1 point to 3rd place
  - Calculated total points for each model at each forecast horizon (table below)
- Key takeaways for this forecasting date
  - LANL performed in the top 3 for most states for 1 week ahead forecasts
  - USC performed well for medium term horizon

Ranking	1 Week Ahead	2 Weeks Ahead	3 Weeks Ahead	4 Weeks Ahead
1	LANL (84)	USC (35)	USC (31)	Karlen (42)
2	JHU-APL (34)	ESG (30)	IEM (30)	Bpagano (33)
3	LNQ (33)	IEM (26)	LNQ (25)	MIT-ISOLAT (30)

Next Steps

- Summarize and analyze model performance by state, forecasting horizon, and model type/methods over multiple forecasting start dates
- Create time series plots of model performance
  - Focus on how model performance changes before and after surges
- CDC changed data structure after December 21st
-

Appendix

Information Spreadsheet

https://docs.google.com/spreadsheets/d/1IWj3ecKVmLZNNWqh3uYqhKuUPbwW2TBvCgOcwc1ZEV4/edit#gid=0