ASKEM 12 Month (July 2023) Ensemble Challenge

Building on the work done for the April ensemble challenge, the goal of the July ensemble challenge is to capture the complexity and nuances around the evolutionary nature of the pandemic. Pretend that you are at various key stages of the pandemic, where there are different dynamics, contexts, and policies in place. Consider the following timepoints during the Covid-19 pandemic:

1. Timepoint 1: April 3rd, 2020, location = New York state, at the beginning of the pandemic, when the main preventive measure was masking. No vaccines were available during this time.

2. Timepoint 2: July 15th, 2021, location = New York state, during the upswing of the Covid wave caused by the arrival of the Delta variant. Vaccines were available at this time.

3. Timepoint 3: January 4th, 2022, location = New York state, when at-home testing was made available, coinciding with the winter Covid wave caused by the original Omicron variant

Bonus: For all 3 timepoints, now consider Texas, which had different Covid-19 dynamics compared to the northeastern coastal states.

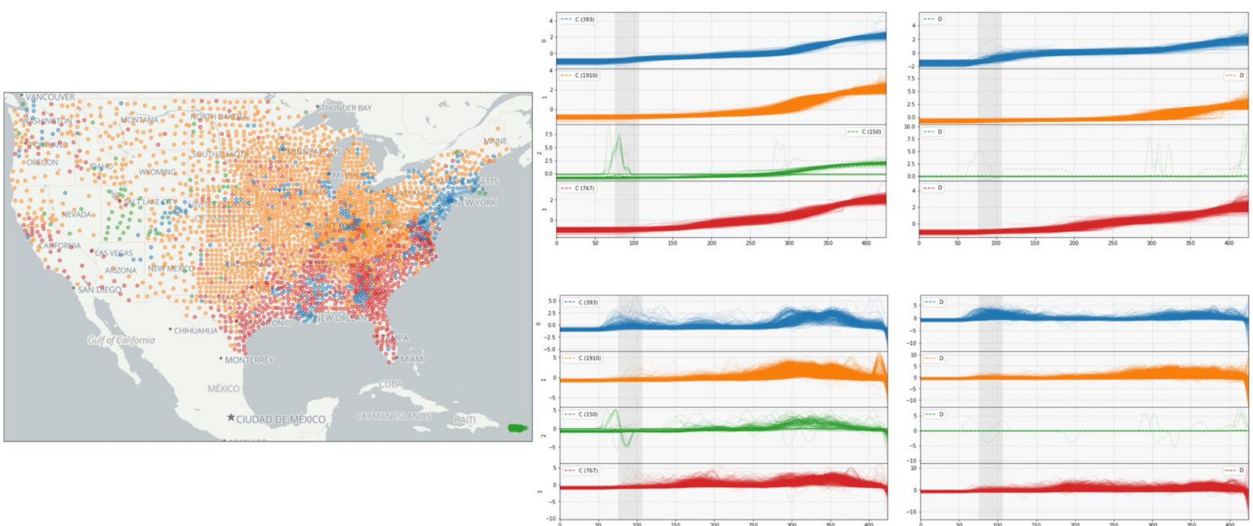

Figure 1. Analysis from ASKE-E that found different Covid-19 dynamics occurring in different regions of the United States.

For each of these timepoints, consider the following:

- What is the most relevant data to use for model calibration?
- What was our understanding of Covid-19 viral mechanisms at the time? E.g. early in the pandemic, we didn't know if reinfection was possible or a common occurrence.
- What are the values of parameter values related to contagiousness/transmissibility, and severity of the dominant strain at the time?

- What policies were in place for a stated location, and how can this information be incorporated into models? There are several databases that track time series of interventions, for example the Oxford COVID-19 Government Response Tracker.

For each timepoint:

1. Show that you can take a single model, calibrate it using any historical data prior to the given date, and predict cases, hospitalizations, and deaths over a 4-week timeframe beginning from the given date, as done in the CDC forecasting challenge. Evaluate the forecast using the Covid-19 forecasting hub error metrics (WIS, MAE) described below. The single model forecast should be evaluated in the same way as the ensemble.
2. Do the same thing with an ensemble of different models.
    a. For the ensemble, it is fine to calibrate each model independently and weight them naively based on their skill.
    b. It would also be fine to calibrate the ensemble as a whole, assigning weights to the different component models, so that you minimize the error of the ensemble on the historical data.
    c. Use the calibration scores and error metrics computed by the CDC forecasting hub. As stated on the website:
    "Periodically, we evaluate the accuracy and precision of the ensemble forecast and component models over recent and historical forecasting periods. Models forecasting incident hospitalizations at a national and state level are evaluated using adjusted relative weighted interval scores (WIS, a measure of distributional accuracy), and adjusted relative mean absolute error (MAE), and calibration scores. Scores are evaluated across weeks, locations, and targets. You can read a paper explaining these procedures in more detail, and look at the most recent monthly evaluation reports. The final report that includes case and death forecast evaluations is 2023-03-13."
3. Produce the forecast outputs in the format specified by the CDC forecasting challenge, including the specified quantiles.

For gold standard data to evaluate your forecasts, use the following sources:

- **Cases**:
    - Johns Hopkins: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_US.csv
    - Reich lab (pulled from Johns Hopkins, but formatted): https://github.com/reichlab/covid19-forecast-hub/blob/master/data-truth/truth-Incident%20Cases.csv ; https://github.com/reichlab/covid19-forecast-hub/blob/master/data-truth/truth-Cumulative%20Cases.csv
- **Hospitalizations**: https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/g62h-syeh
- **Deaths**:

- Johns Hopkins (also contains population by county): https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_US.csv
- Reich lab (pulled from Johns Hopkins, but formatted): https://github.com/reichlab/covid19-forecast-hub/blob/master/data-truth/truth-Incident%20Deaths.csv ; https://github.com/reichlab/covid19-forecast-hub/blob/master/data-truth/truth-Cumulative%20Deaths.csv

Model considerations:

- You may consider any of the models you have seen in the starter kit, or 6-month hackathon and evaluation scenarios
- You may search for new models in the literature, or use TA2 model extension/transformation capabilities to modify models already in Terarium

Bonus Question: Consider a situation where we would want to create an ensemble from various parameter configurations of a single model. From November 7th, 2022, the United States started transitioning from BA.5 as the dominant Covid-19 strain, to BQ.1 as the dominant strain. Develop an ensemble model using data over a 1-month period starting from November 7th, using different parameterizations of a single model, reflecting the different competing strains during this period.