

# COVID-19 Forecast Similarity Analysis

Johannes Bracher, Aaron Gerding, Evan Ray, Nick Reich, Nutch Wattanachit

07/07/2021

## Forecast inclusion criteria

- Models: All models with complete submissions for the following criteria
- Targets: 1-4 wk ahead inc death and inc case
- Target end dates: Oct 19th, 2020 - May 24th, 2021
- Probability levels: All
- Locations:
  - 5 states with highest cumulative deaths by February 27th, 2021: CA, FL, NY, PA, TX
  - 5 states with highest cumulative cases by February 27th, 2021: CA, FL, IL, NY, TX
  - 5 states with lowest cumulative deaths by February 27th, 2021: AK, HI, ME, VT, WY
  - 5 states with lowest cumulative cases by February 27th, 2021: DC, HI, ME, VT, WY

## 1-4 Week Ahead Incident Death Forecasts

Naturally, the differences between the two approximations are larger for further horizons since forecasts are more dissimilar. The differences for the approx. CD between CU-select and the ensemble forecasts seem a bit more pronounced for all horizons - we might want to check how the CDF (built from quantiles) look.

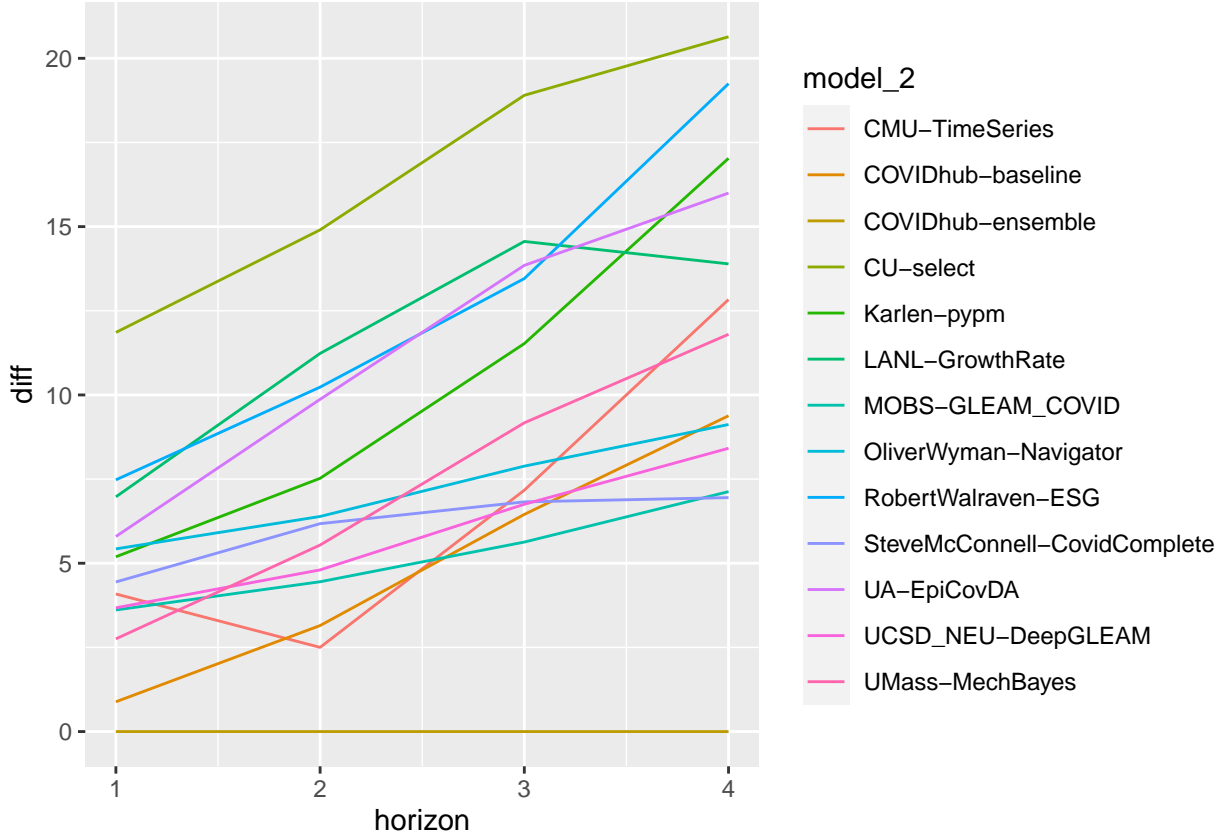
There are 13 models that fulfilled the criteria for the 5 locations with highest cumulative deaths.

## Model types

Model	Type
CMU-TimeSeries	statistical
COVIDhub-baseline	statistical
COVIDhub-ensemble	ensemble
CU-select	mechanistic
Karlen-pypm	mechanistic
LANL-GrowthRate	statistical
MOBS-GLEAM_COVID	mechanistic
OliverWyman-Navigator	mechanistic
RobertWalraven-ESG	statistical
SteveMcConnell-CovidComplete	statistical
UA-EpiCovDA	mechanistic
UCSD_NEU-DeepGLEAM	mechanistic
UMass-MechBayes	mechanistic

## Differences between two approximations (for high count locations only)

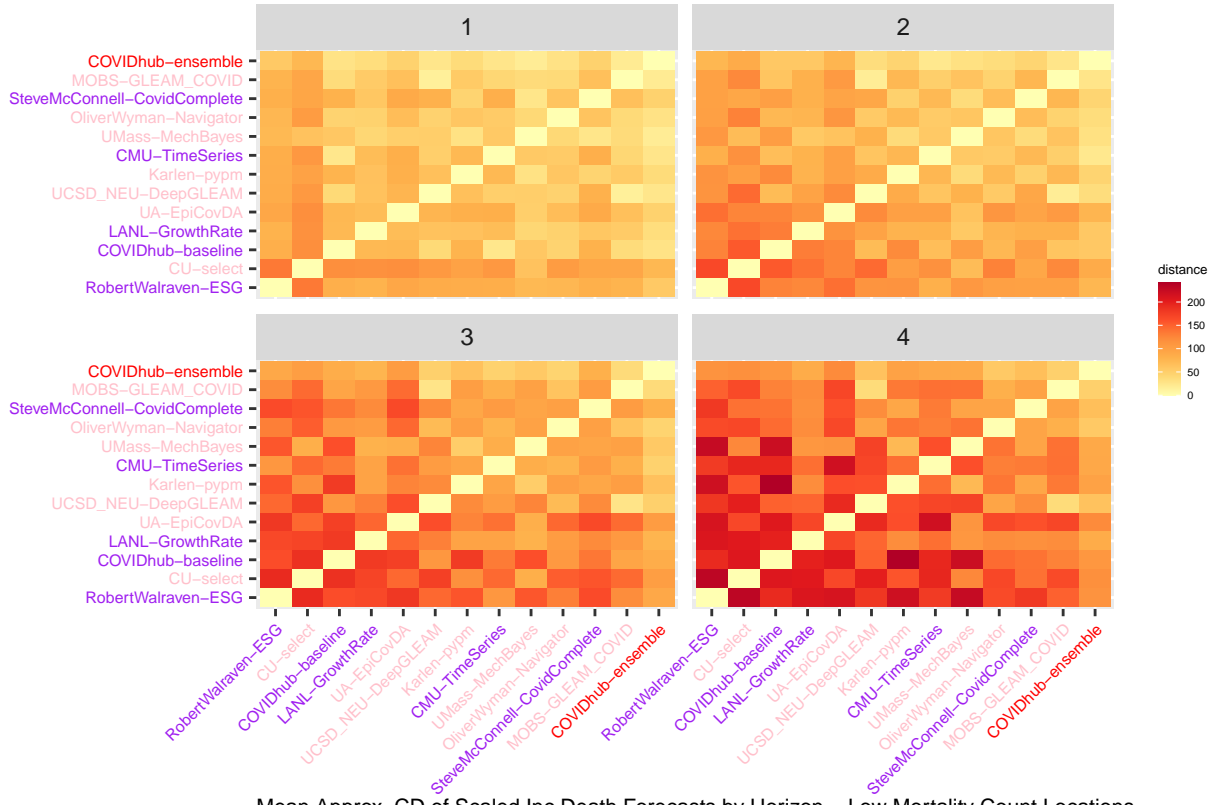
The approximated pairwise Cramér’s distances between each forecast and the ensemble are calculated using both types of approximations to check for any large discrepancies between the two methods. The table below shows the averaged approx. CD over all target end dates and all 5 high count locations. Given the differences are positive (unequal-equal) and the equally-spaced formula overweighs the tail quantiles (it actually distort the height of all boxes since it assumes  $1/23$  (0.043) increment here when most increments are 0.05), the tail differences might be offset by the underestimation of the heights at most quantiles by the equally-weighted formula.



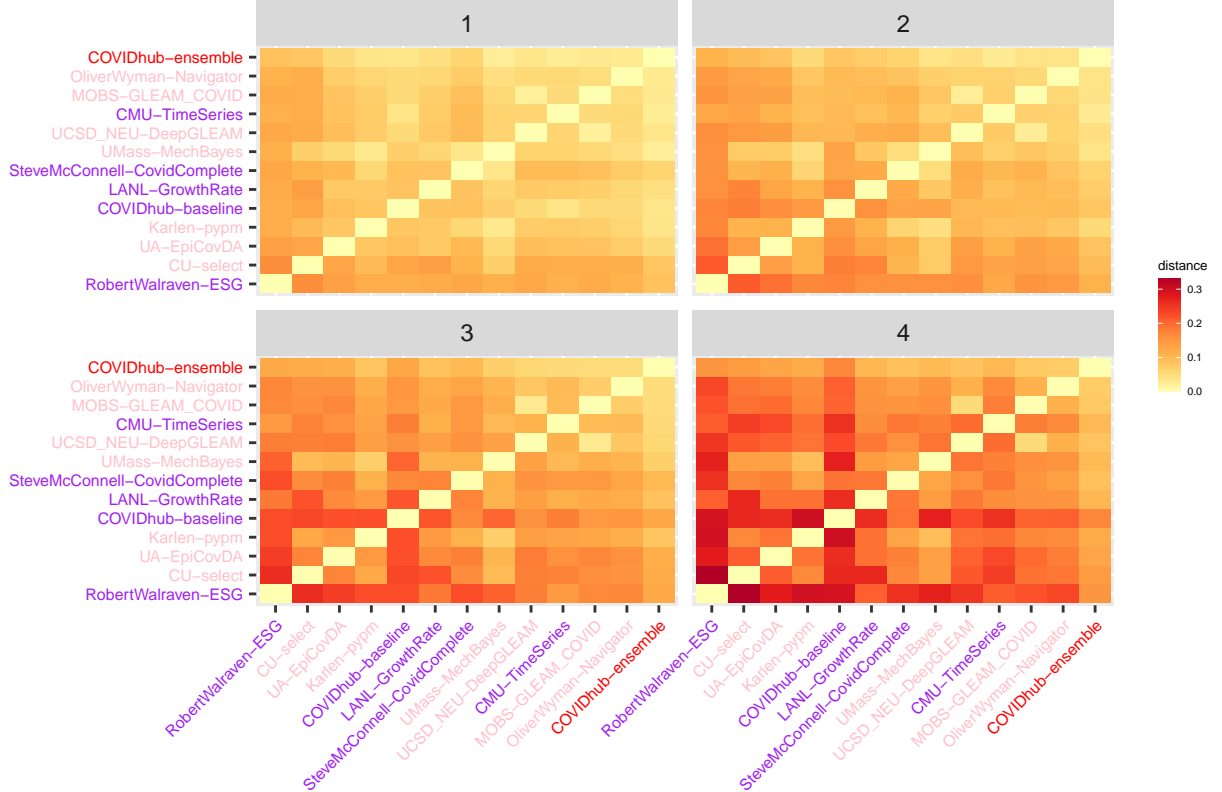
## Mean approximated pairwise distances over across 5 high count locations

We can visualize the mean approximated pairwise distances across all weeks and locations in heat maps. The distance from the model to itself is zero. The  $x$ -axis is arranged based in an ascending order of the model’s approximate pairwise distance from the COVIDhub-ensemble. So, the first model is the model that is most dissimilar (on average) to the ensemble in this time frame.

Mean Approx. CD of Inc Death Forecasts by Horizon – High Mortality Count Locations



Mean Approx. CD of Scaled Inc Death Forecasts by Horizon – Low Mortality Count Locations

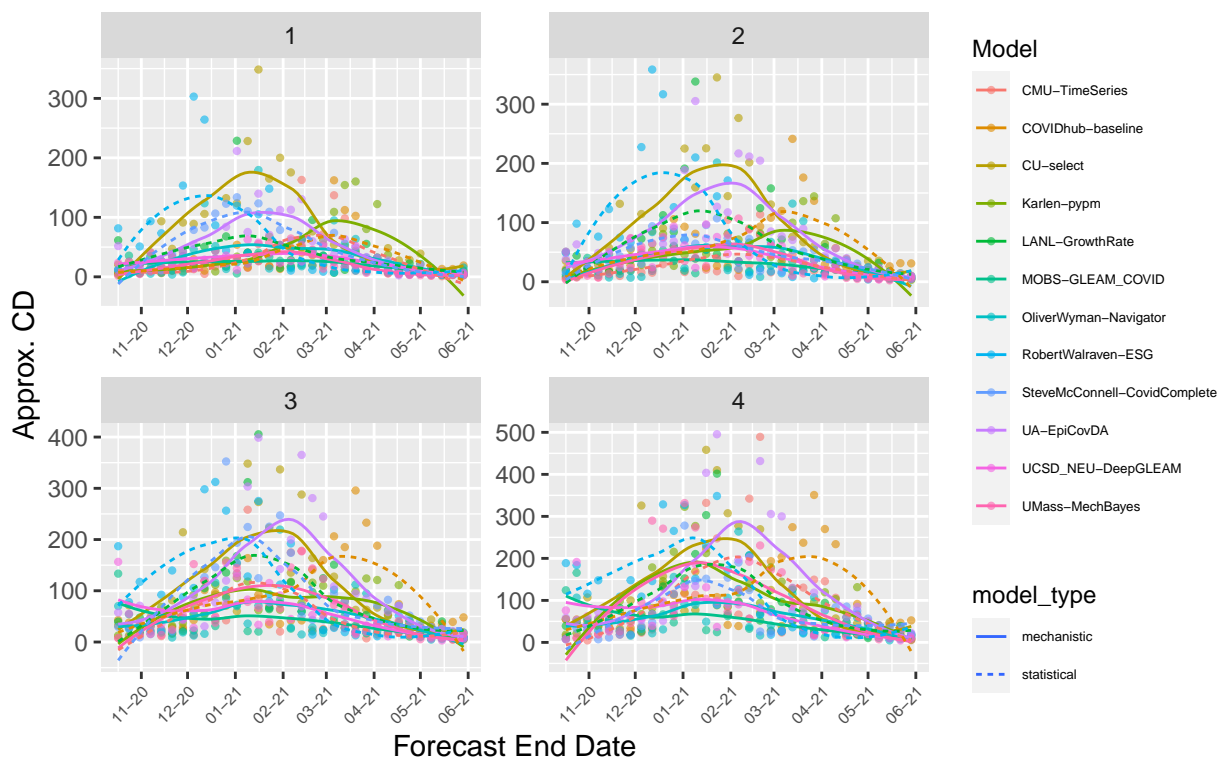


For high mortality count locations, the distances between pairs of forecasts are higher for further forecast

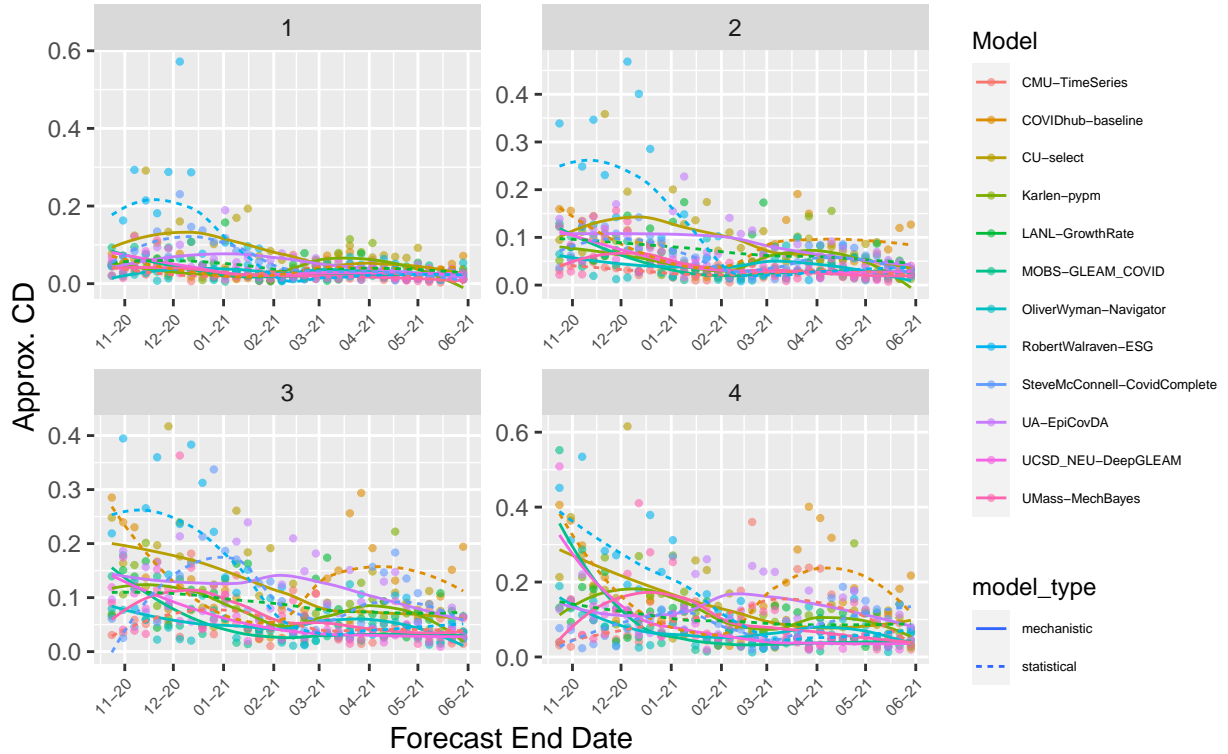
horizons. TCU-select forecasts for high count locations seem to be more dissimilar to other models on average across all forecast horizons. After scaling by the truth at t-1, some distances get “smoothed out”, but the observed trends persist.

We can also look at the mean approximated pairwise distances across locations only to see how the models become more similar or dissimilar over time.

### Mean Approx. CD from COVIDhub–ensemble Over Time – High Mortality Count Locations



## Mean Approx. CD (Scaled) from COVIDhub-ensemble Over Time – High Mortality Count Locations

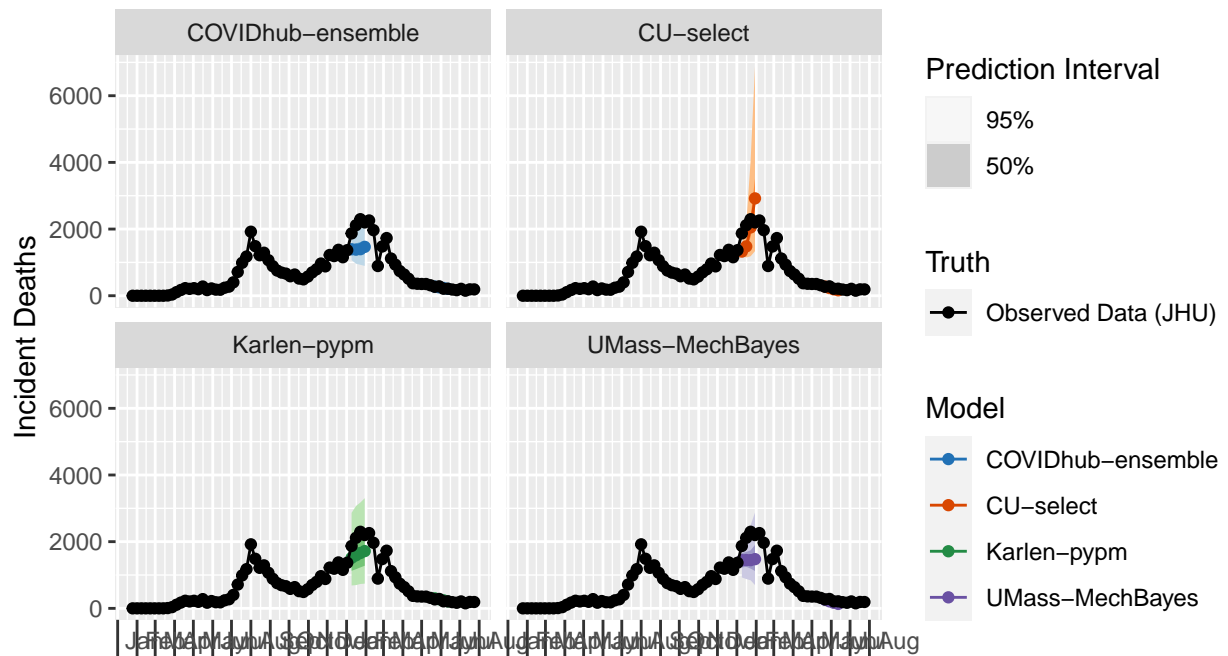


Below is the plots of forecasts in the forecast week where the mean WIS for the ensemble is the highest (for now just Texas). 1-4 wk ahead from this forecast date is from early Jan to early Feb target end dates (which is when, on average across all locations, distance from the ensemble is high).

## Weekly COVID-19 Incident Deaths: observed and forecasted

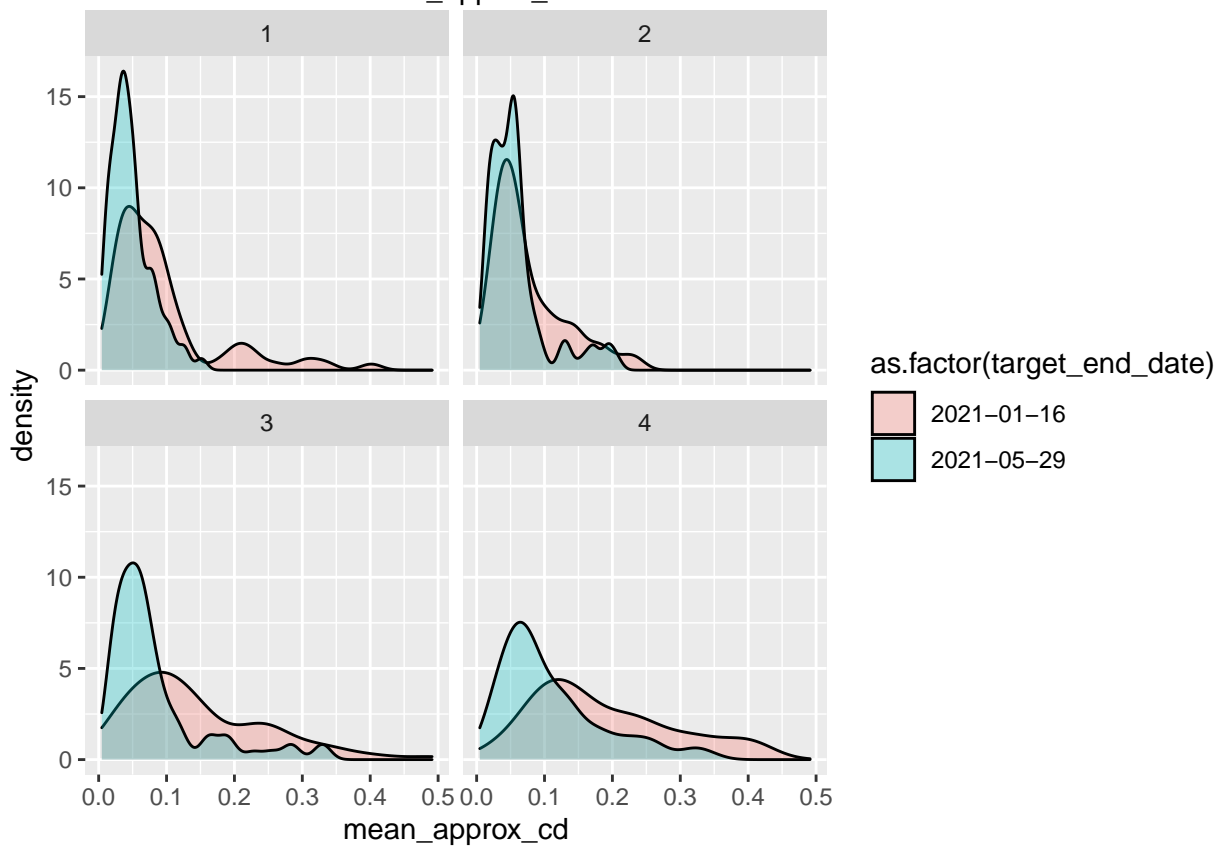
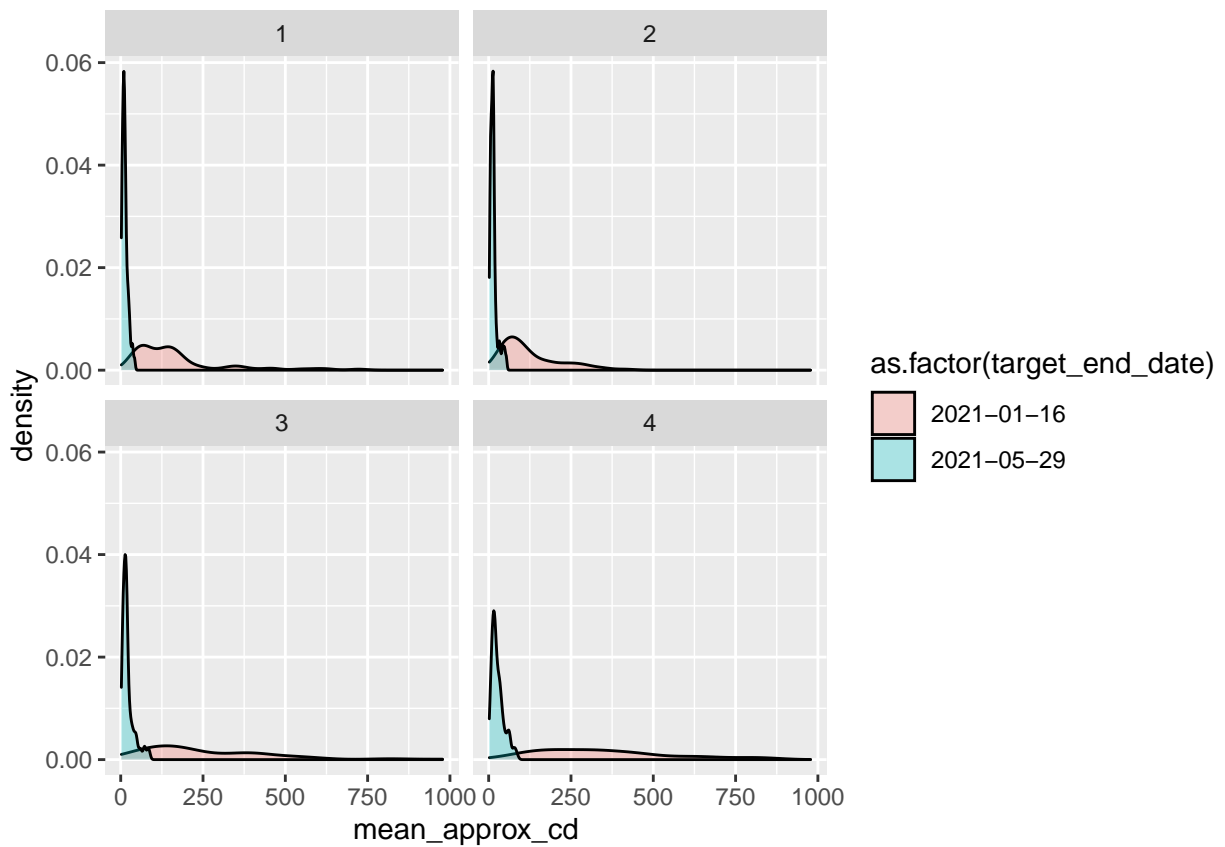
Selected location(s): Texas

Selected forecast date(s): 2021-01-04, 2021-01-03, 2021-05-17, 2021-05-16



i), COVIDhub-ensemble, UMass-MechBayes, CU-select, Karlen-pypm (forecasts)

We can check the distribution of all mean pairwise distances across all locations in that week:



## Hierarchical clustering based on mean approx. CD across all weeks and locations

We can cluster the distances using hierarchical clustering. Different linkages will result in different clusters - here we use ward linkage.

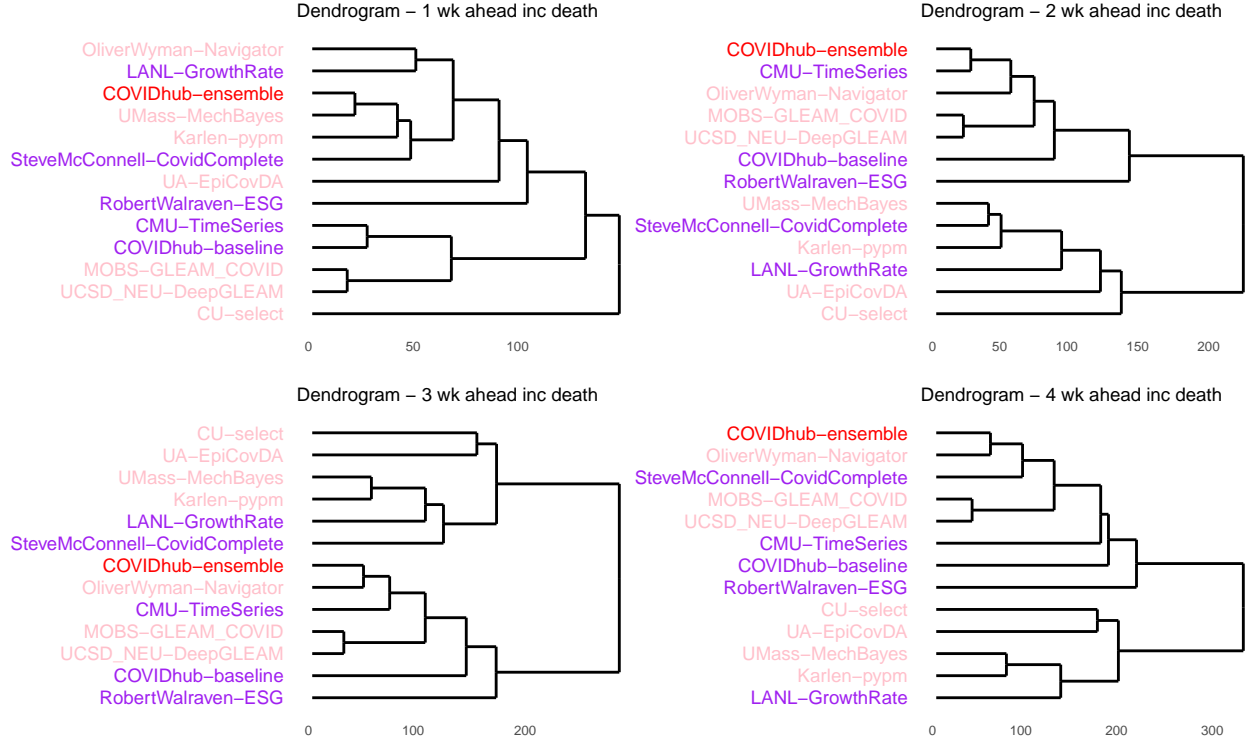
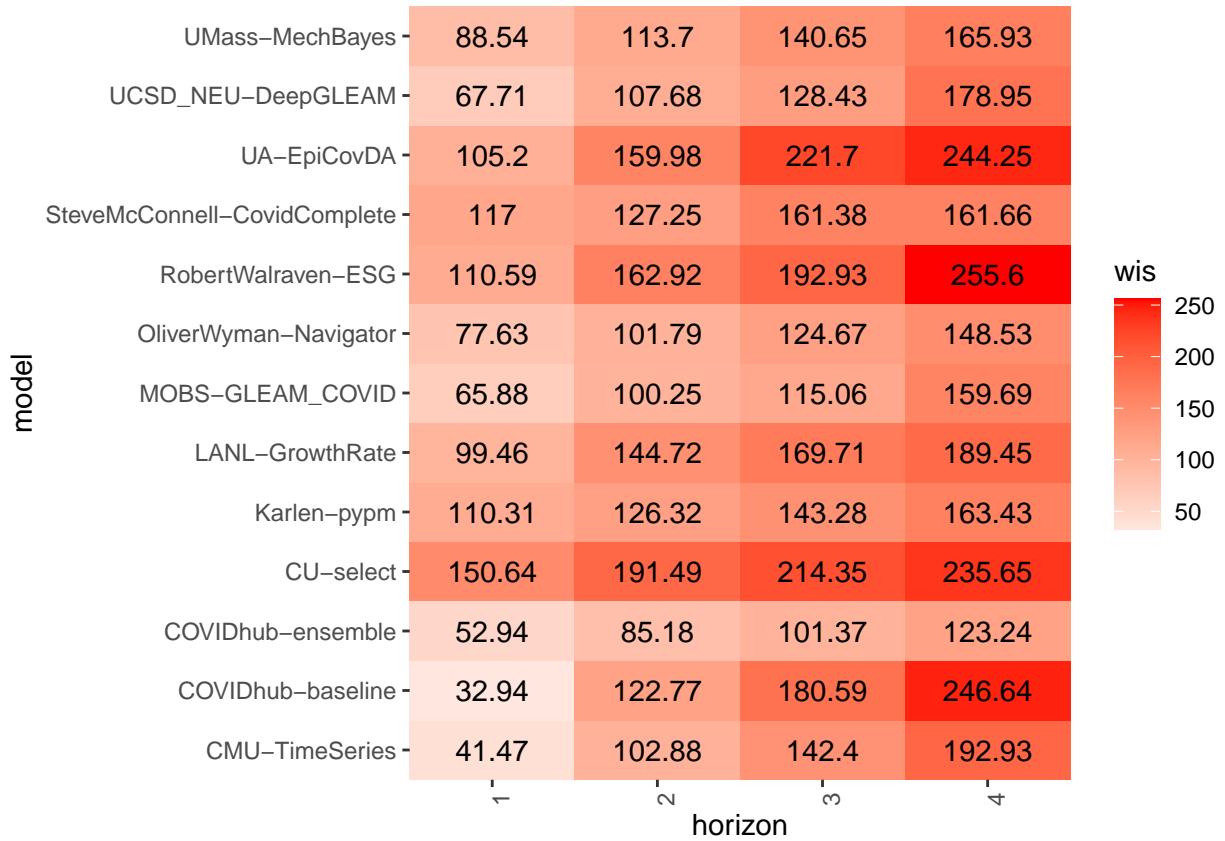


Figure 1: High Mortality Count Locations

We can also check the mean WIS across these location for each horizon. We can see that models with closer WIS are grouped together (CD technically generalizes to WIS in the case where  $G$  is a point mass, so it is expected).





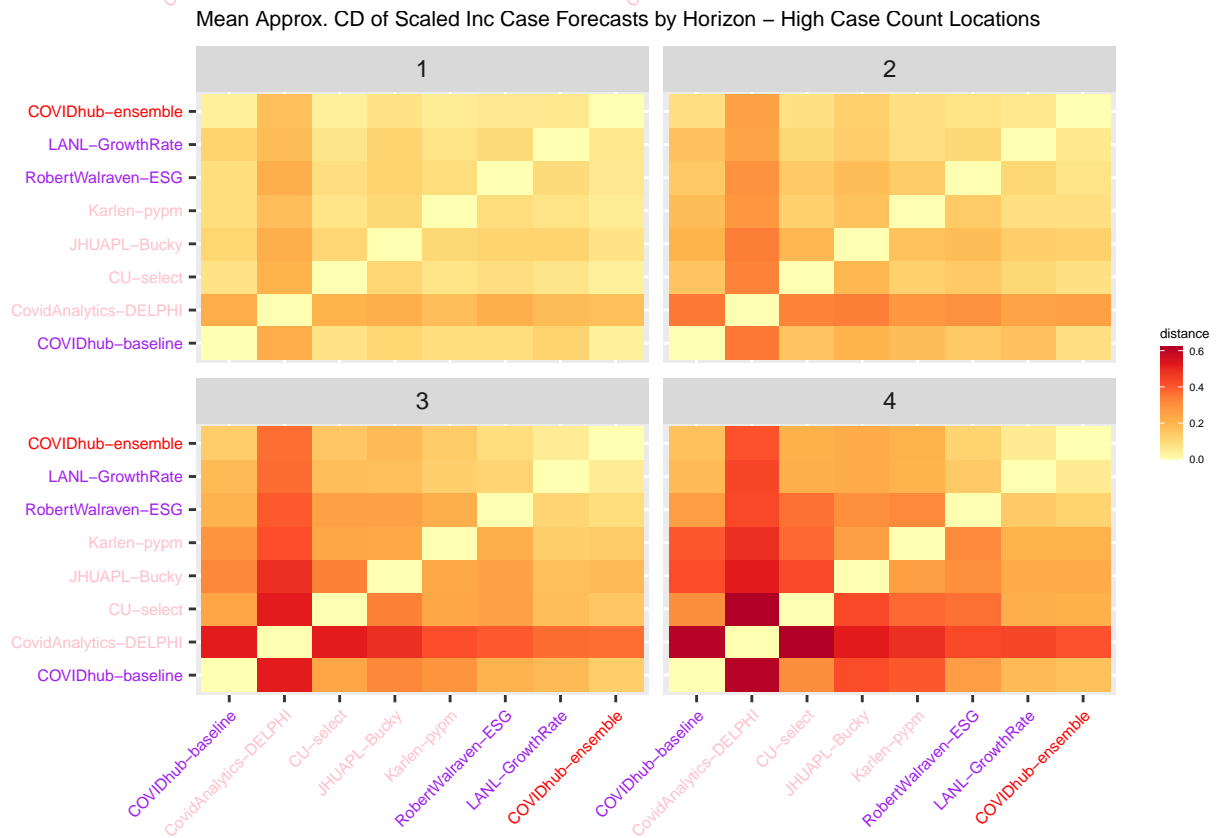
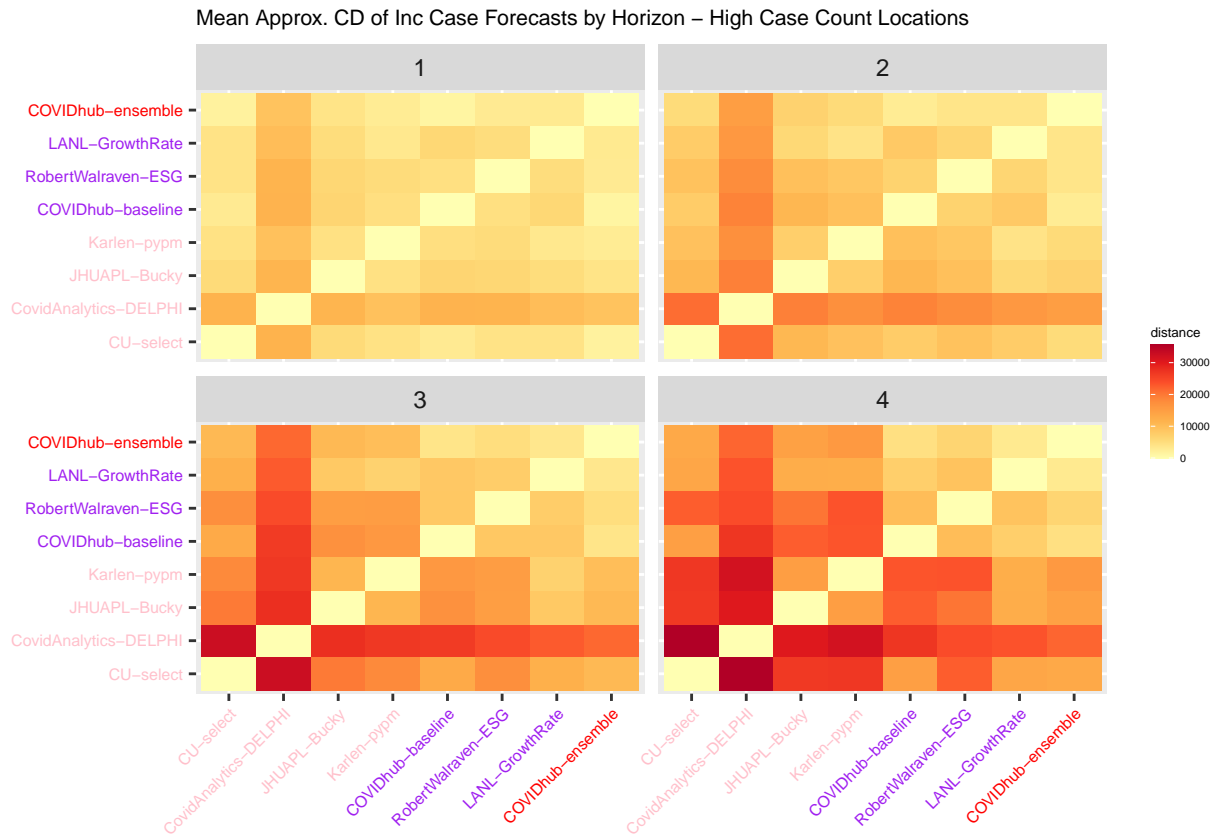
## 1-4 Week Ahead Incident Case Forecasts

### Model types

Model	Type
CovidAnalytics-DELPHI	mechanistic
COVIDhub-baseline	statistical
COVIDhub-ensemble	ensemble
CU-select	mechanistic
JHUAPL-Bucky	mechanistic
Karlen-pypm	mechanistic
LANL-GrowthRate	statistical
RobertWalraven-ESG	statistical



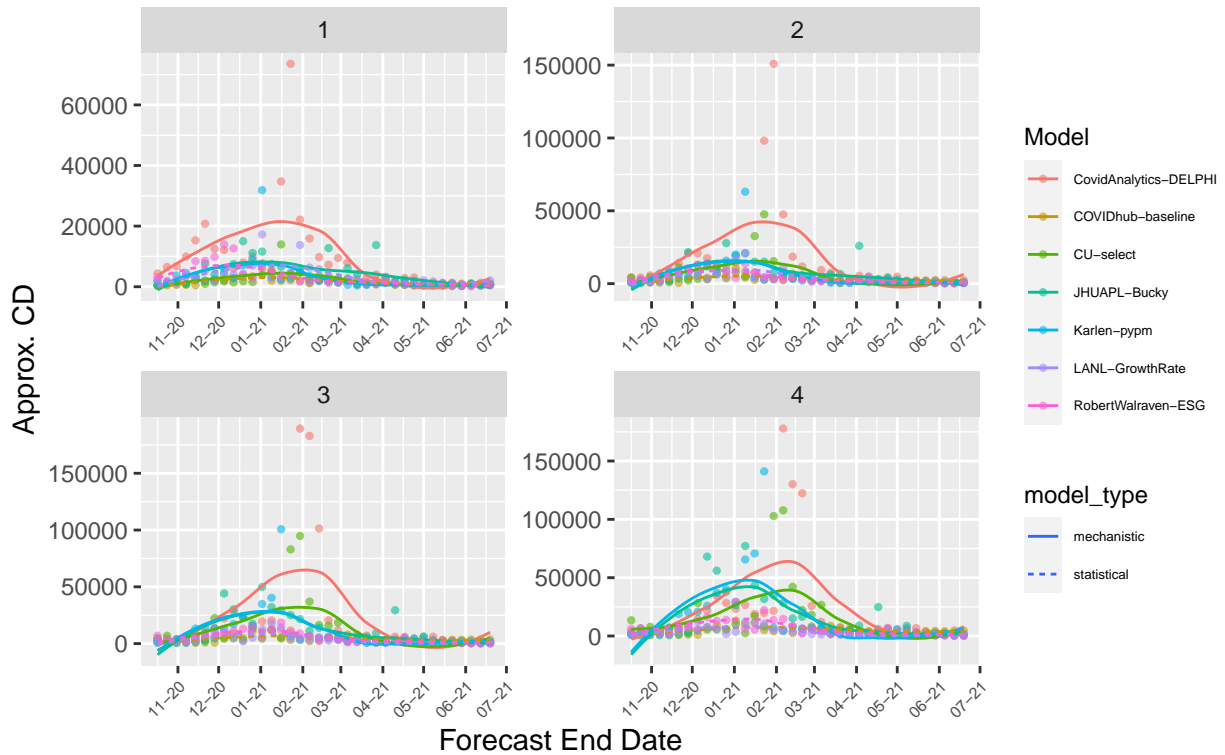
## Mean approximated pairwise distances over across 5 high count and 5 low count locations



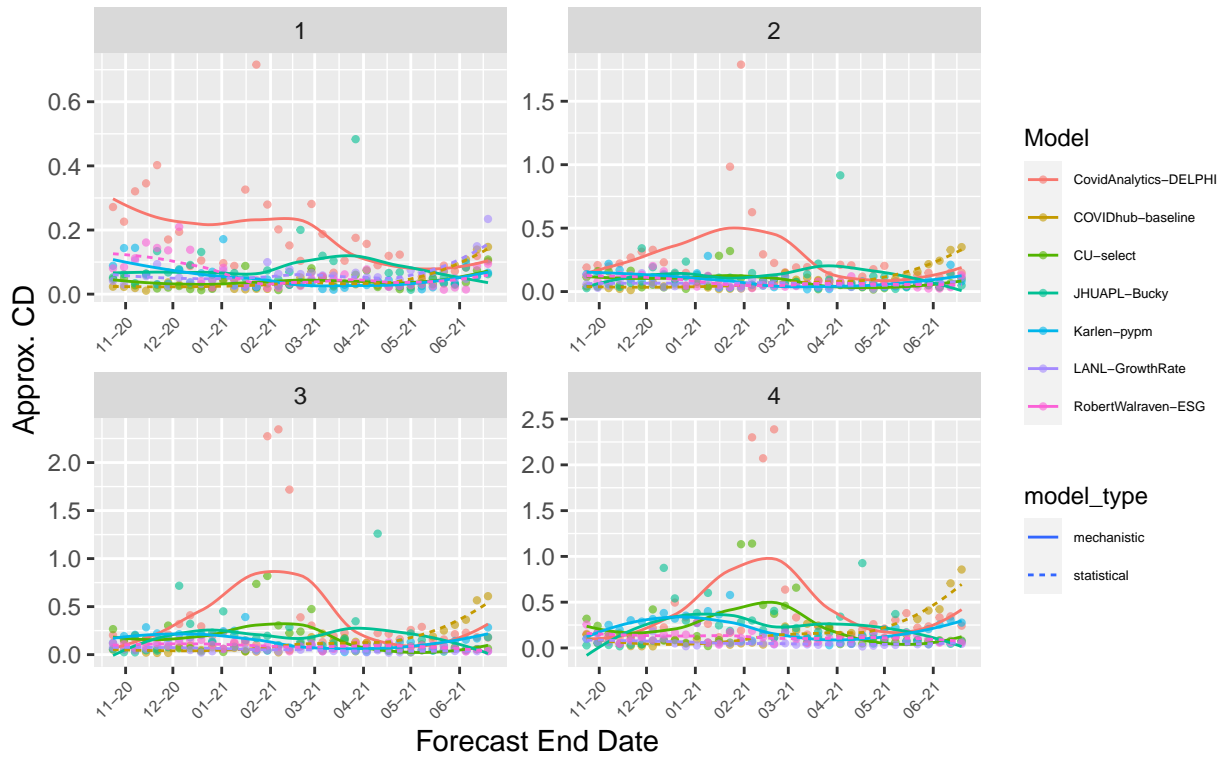
CovidAnalytics–DELPHI forecasts for both high and low count locations seem to be more dissimilar to other models on average across all forecast horizons.

When we look at the approximated pairwise distances over time, we see high distances from the ensemble around Jan-Feb 2021 for high count locations.

### Mean Approx. CD from COVIDhub–ensemble Over Time – High Mortality Count Locations



## Mean Approx. CD (Scaled) from COVIDhub-ensemble Over Time – High Mortality Count Locations

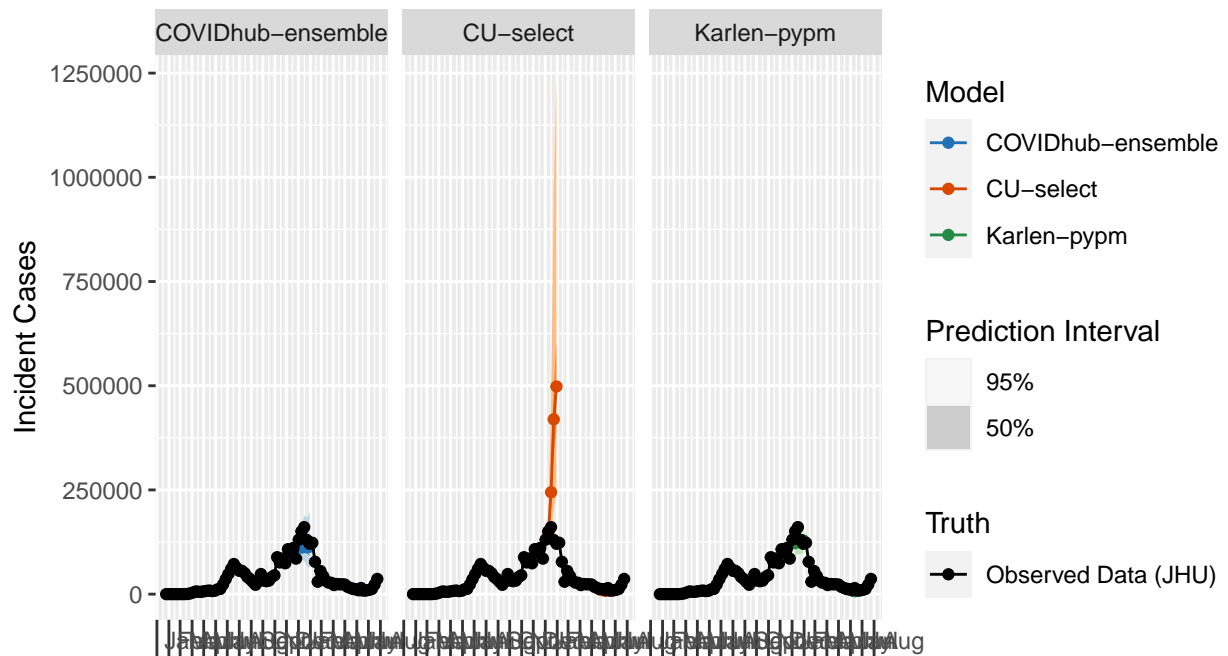


Below is the plots of forecasts in the forecast week where the mean WIS for the ensemble is the highest (for now just Texas). 1-4 wk ahead from this forecast date is from early Jan to early Feb target end dates (which is when, on average across all locations, distance from the ensemble is high).

## Weekly COVID-19 Incident Cases: observed and forecasted

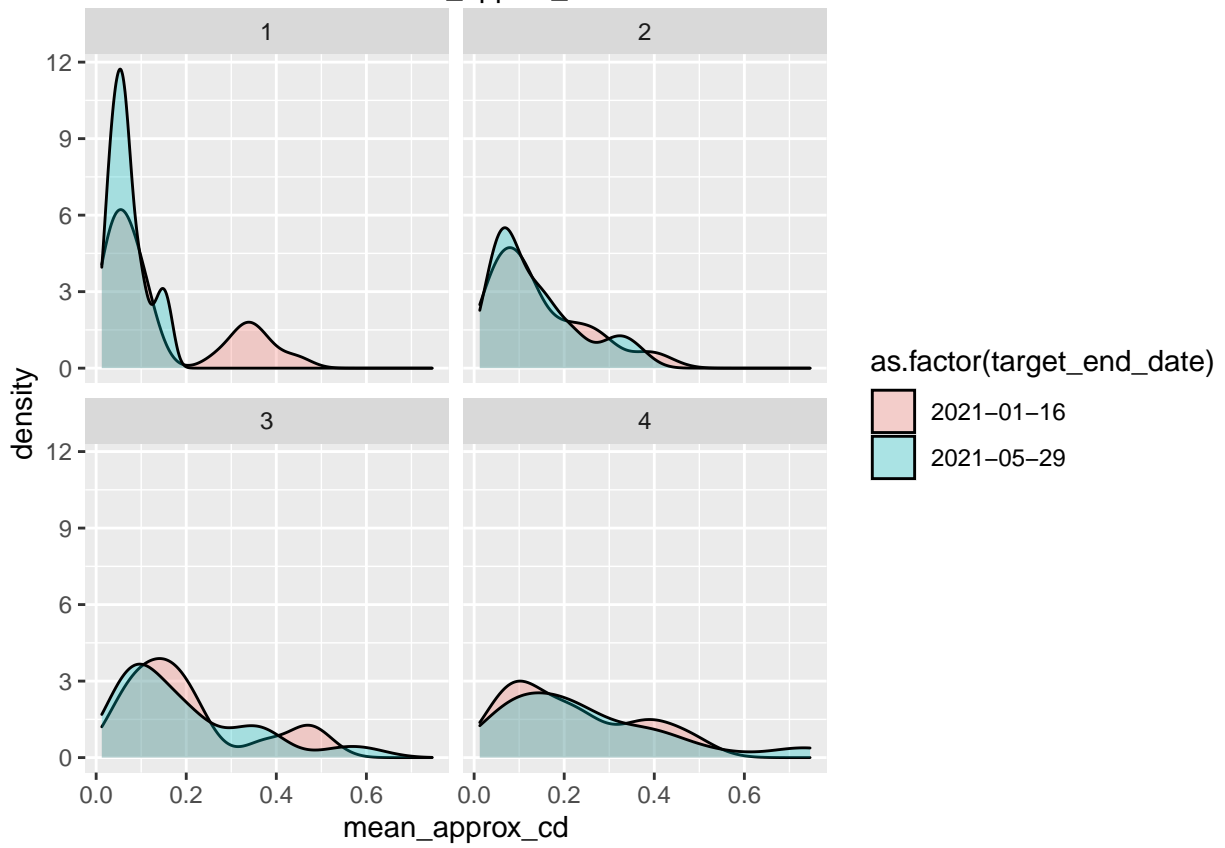
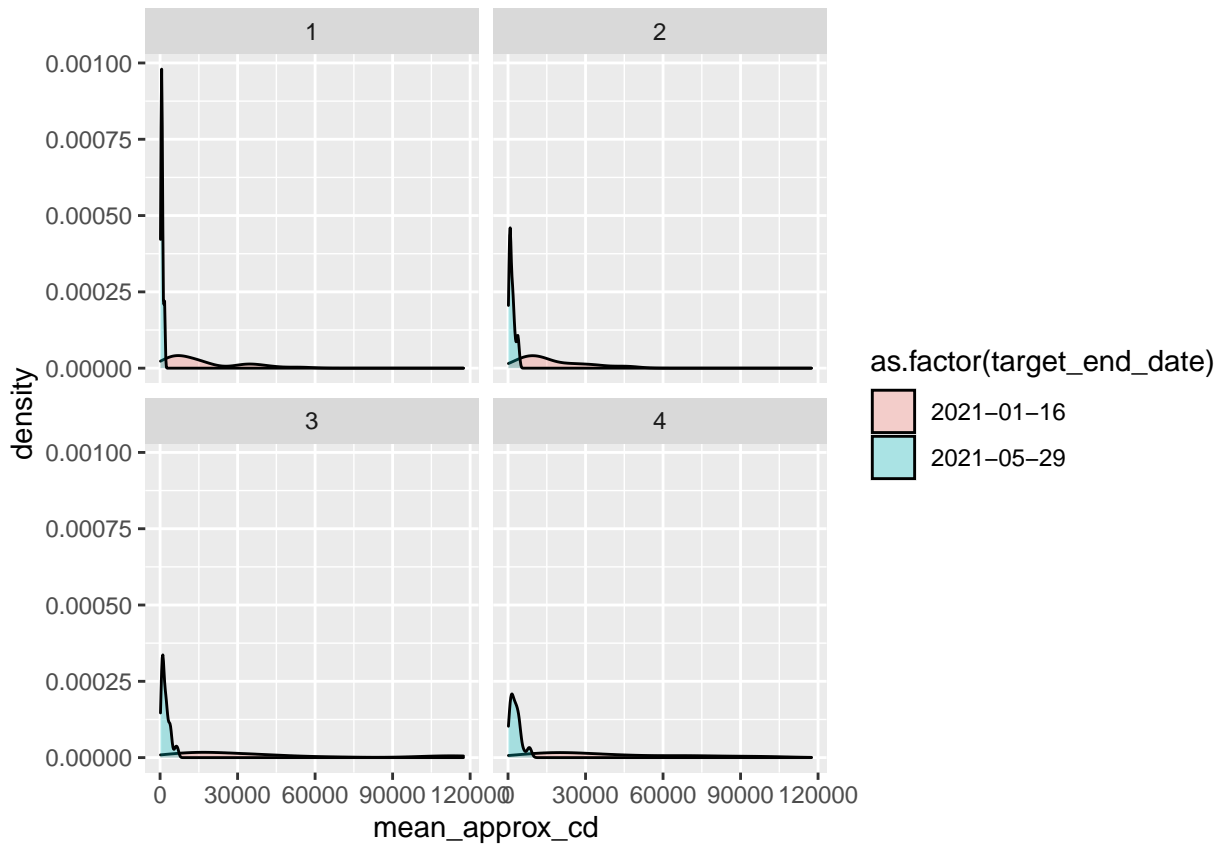
Selected location(s): Texas

Selected forecast date(s): 2021-01-04, 2021-01-03, 2021-05-17, 2021-05-16



:: JHU (observed data), COVIDhub-ensemble, CU-select, Karlen-pypm (forecasts)

We can check the distribution of all mean pairwise distances across all locations in that week:



## Hierarchical clustering based on mean approx. CD across all weeks and locations

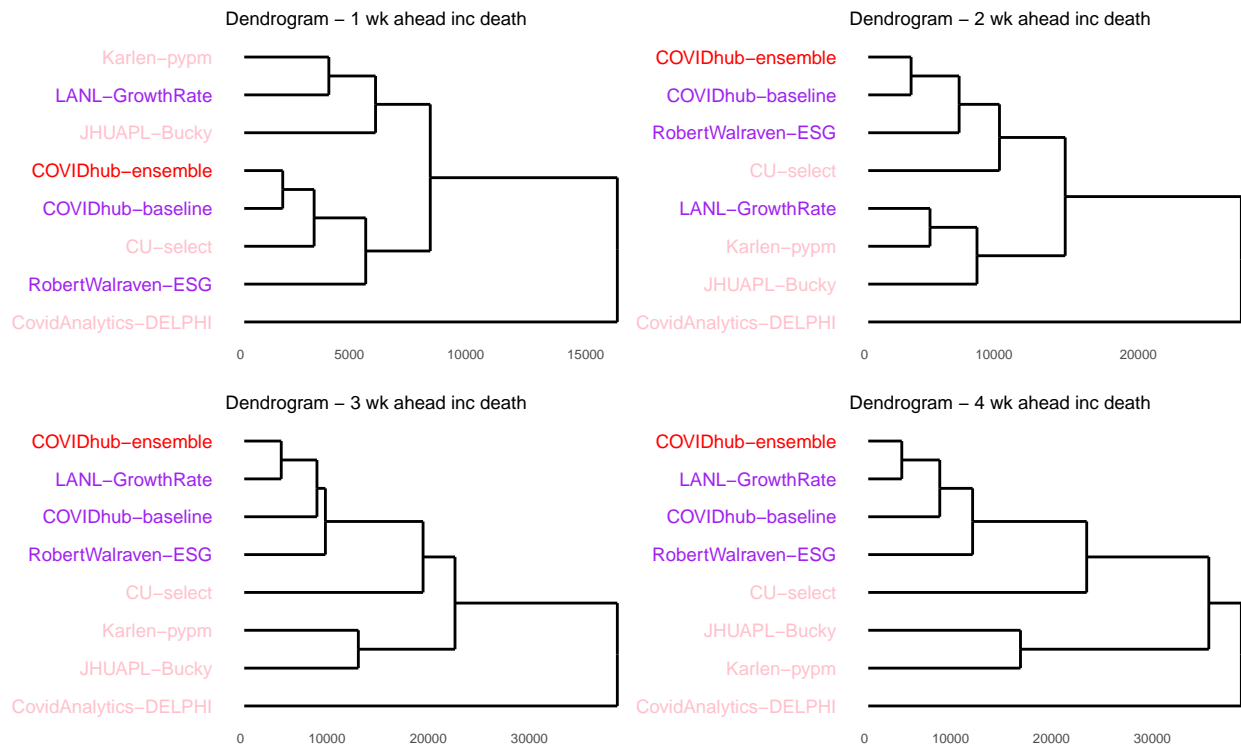


Figure 2: High Case Count Locations

We see similar trend here as we saw in inc death analysis.



