COVID-19 Forecast Similarity Analysis

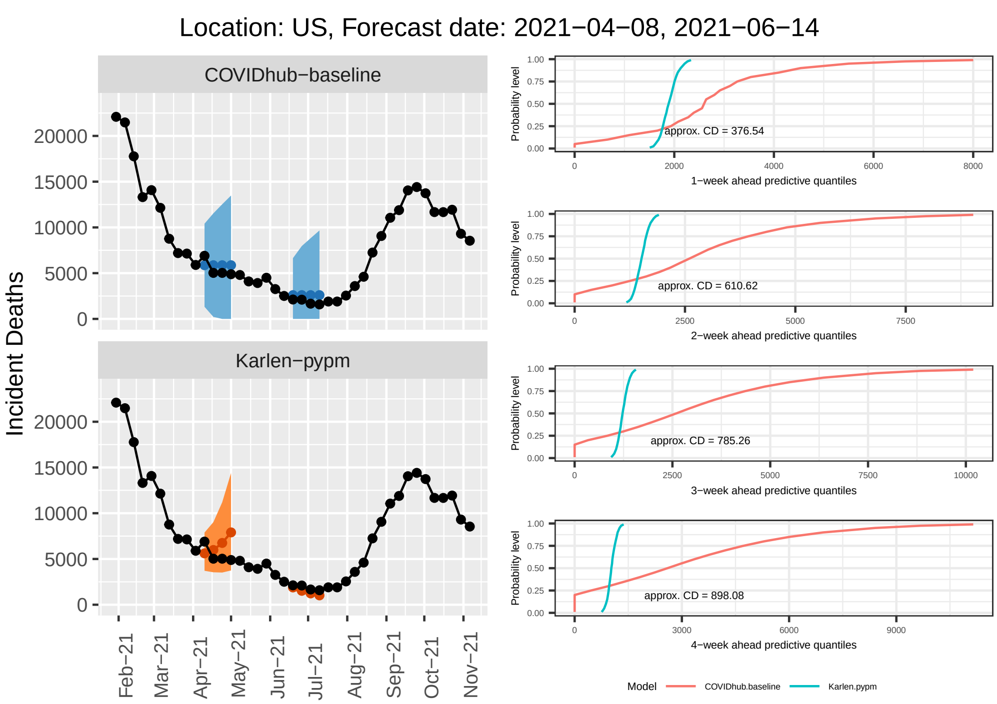
Johannes Bracher, Evan Ray, Nick Reich, Nutcha Wattanachit

03/10/2022

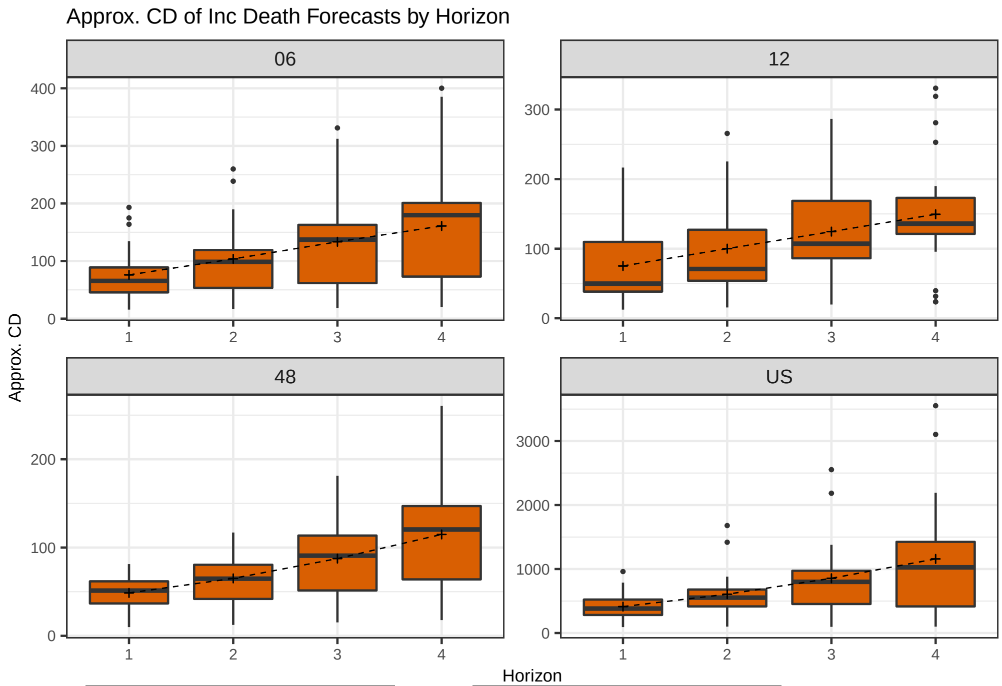
Forecast inclusion criteria

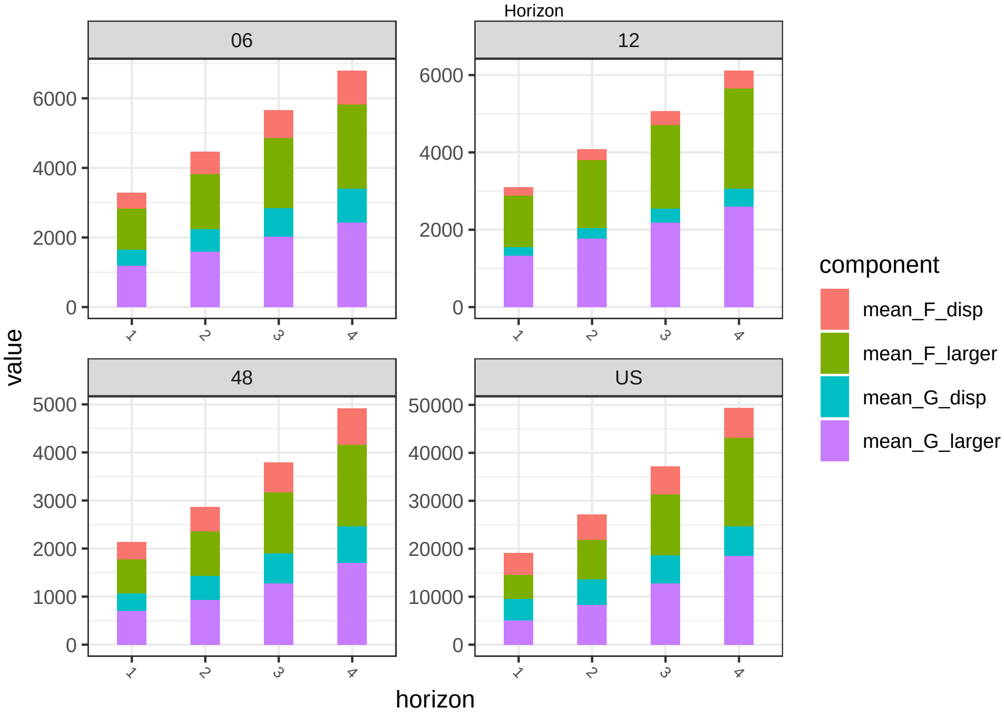
- Targets: 1-4 wk ahead inc death
- Target End Dates: varying sets depending on the analyses
- Probability levels: All
- Locations: 3 states with large cumulative deaths by target end date Feb 5th, 2022 and US National: FL, TX, CA, US

Examples of forecasts and Cramer distances



Overall similarity by horizon





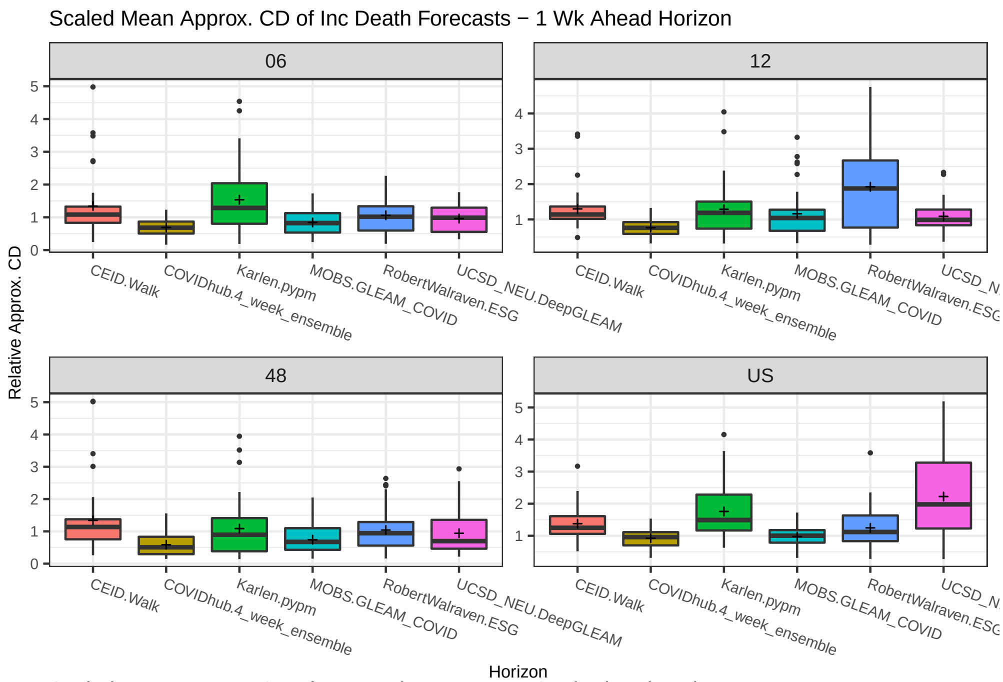
- Most of the distance comes from the the mean across all horizons, except for 1 wk ahead for US.
- Each component looks proportional across horizon, strange?

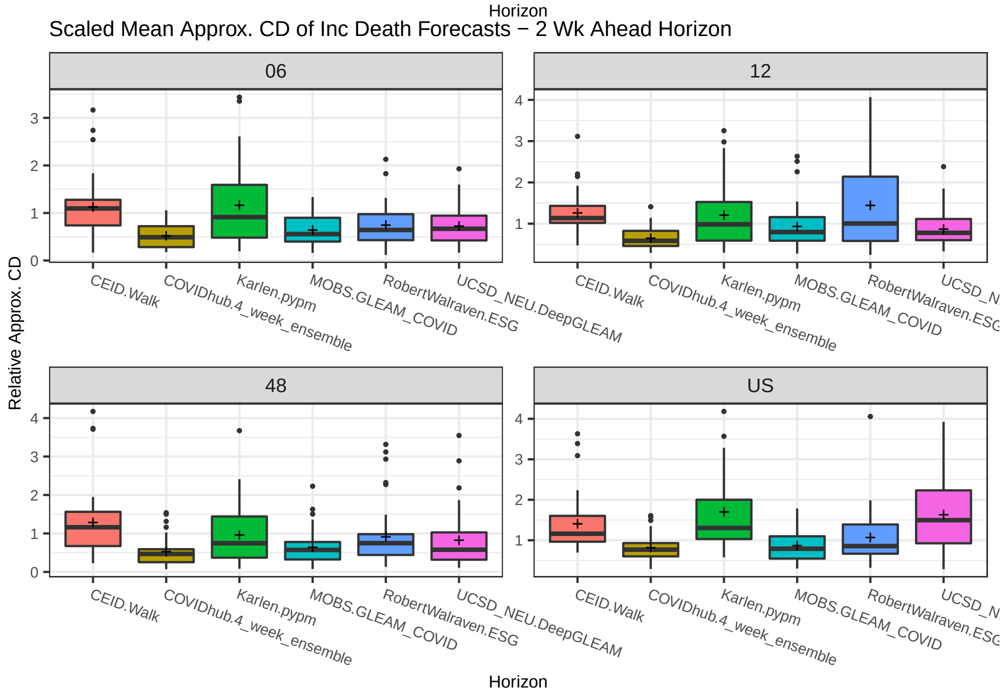
Mean forecast similarity relative to baseline

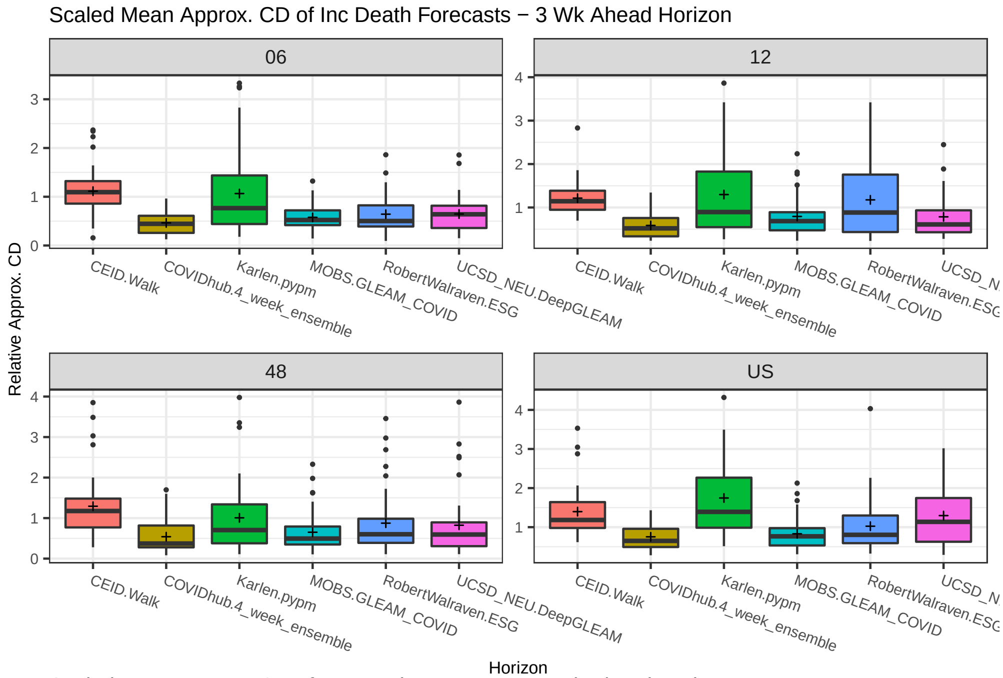
For each model-location-horizon-target_end_date, I calculate the mean of approx. CD of that model to all other models, except baseline (mean of all the pairwise distances with that model as one of the pair).

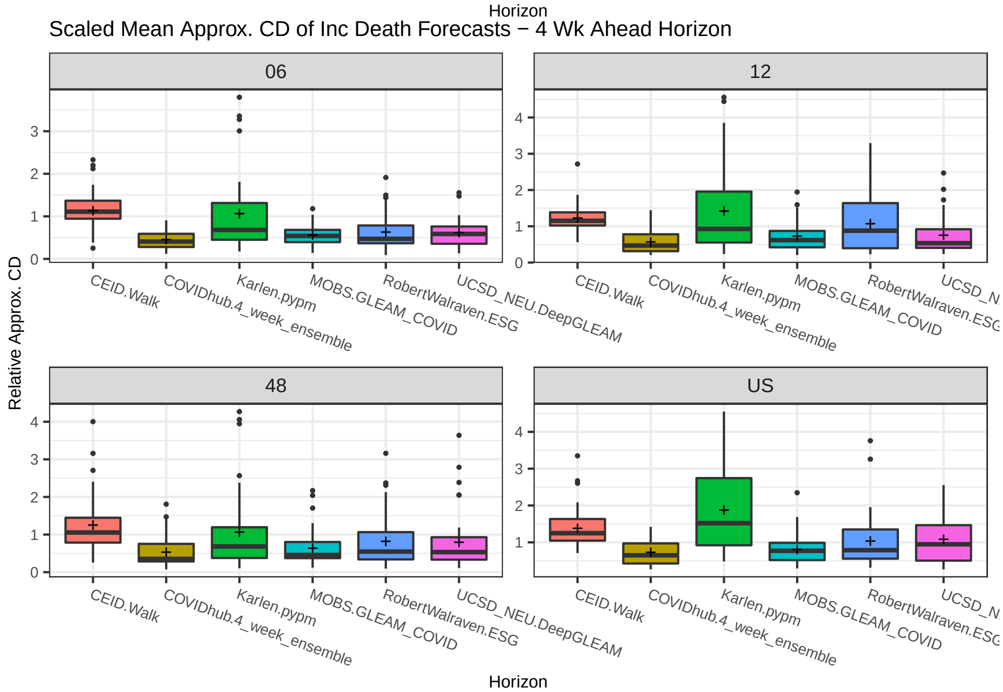
Now we have a mean of CD between one model to all other models for each models, then we scale them by dividing by the mean of CD between baseline to all other models.

The interpretation is a little tricky. - If we see the value around 1, we can say that the mean distance of model A to other models is similar to the mean distance of baseline to other models. - Less than 1 implies the mean distance of model A to other models is less than that of baseline to other models (model A is more similar to other models relative to baseline's similarity to others). - Higher than 1 implies model A is less similar to others relative to baseline's similarity to others.





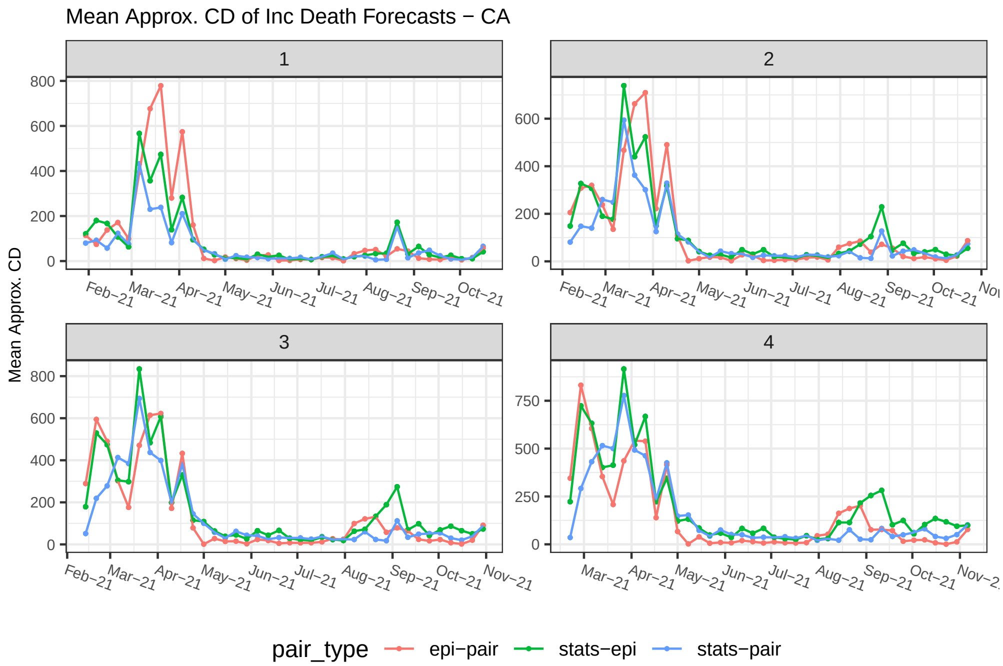


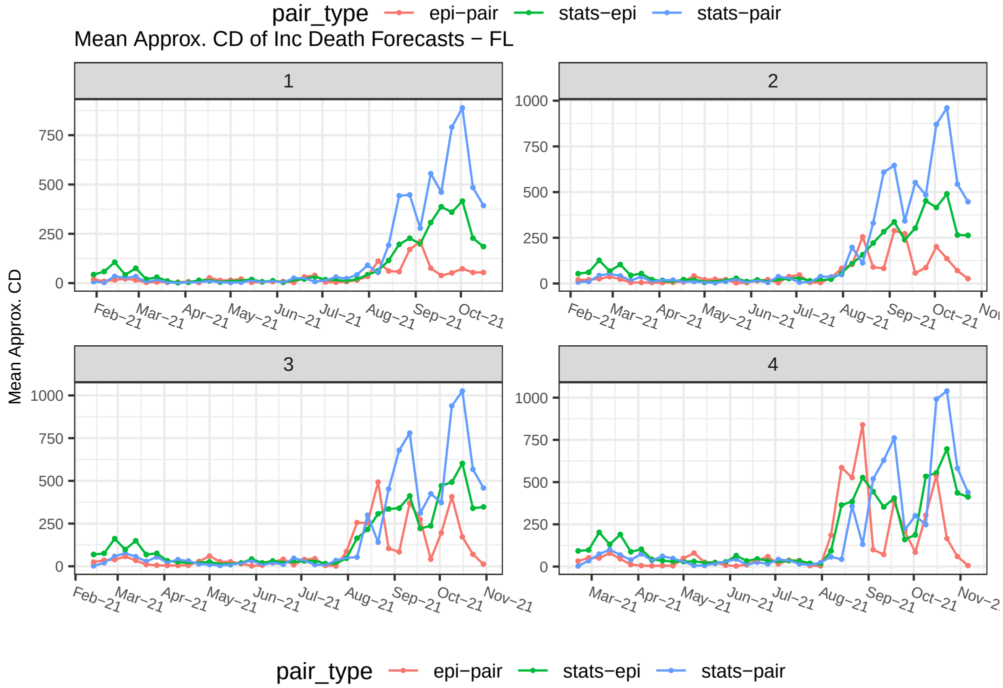


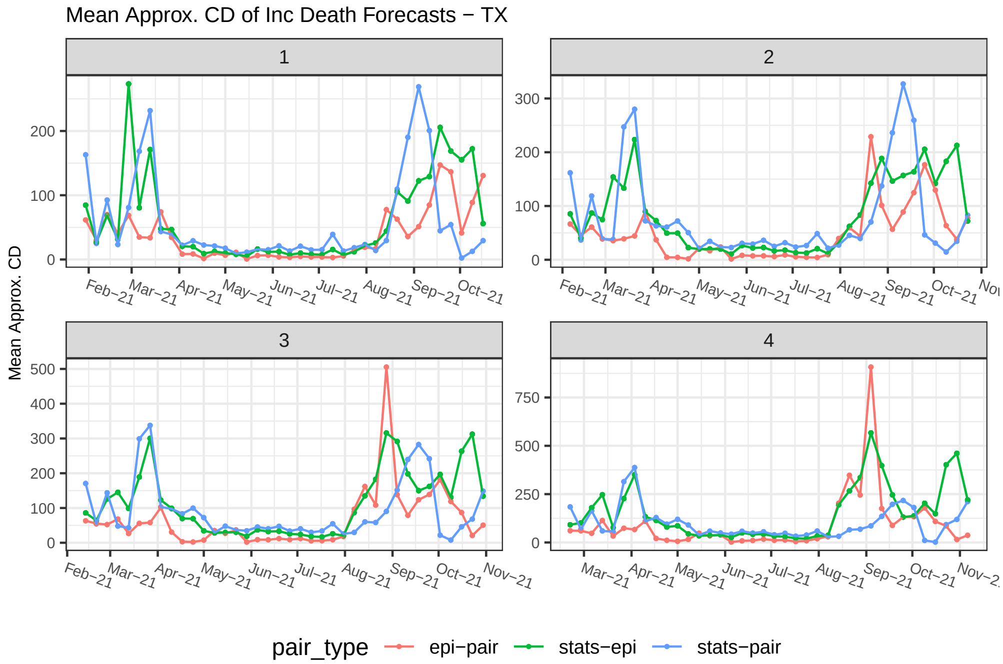
- The results look pretty similar across all horizons, but there are differences across locations in each horizon.
- Seeing CEID-walk around 1 with low-ish variation may be encouraging? It's a non-epi model and the similarity to others is about the same relative to baseline.
- RW-ESG is categorized as stats model, but it is more similar to others relative to baseline (but the range is pretty big).
- There is high variation in the similarity between Karlen-pypm and other models relative to baseline's similarity to others.
- Ensemble is more similar to other models compared to baseline's similarity to other models across all horizon-location pair.

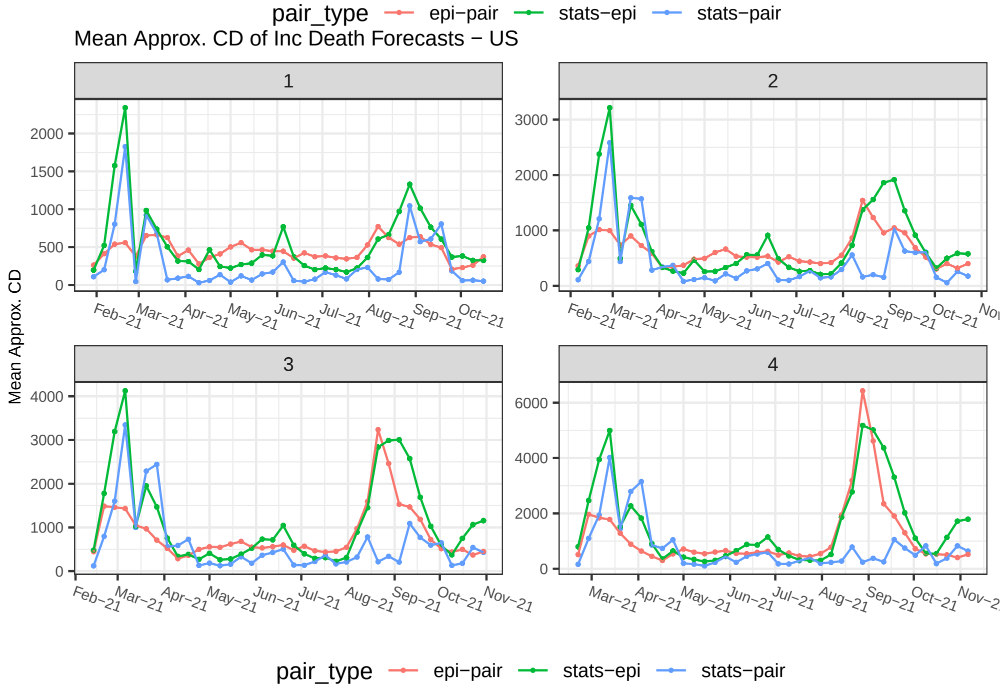
Similarity by type

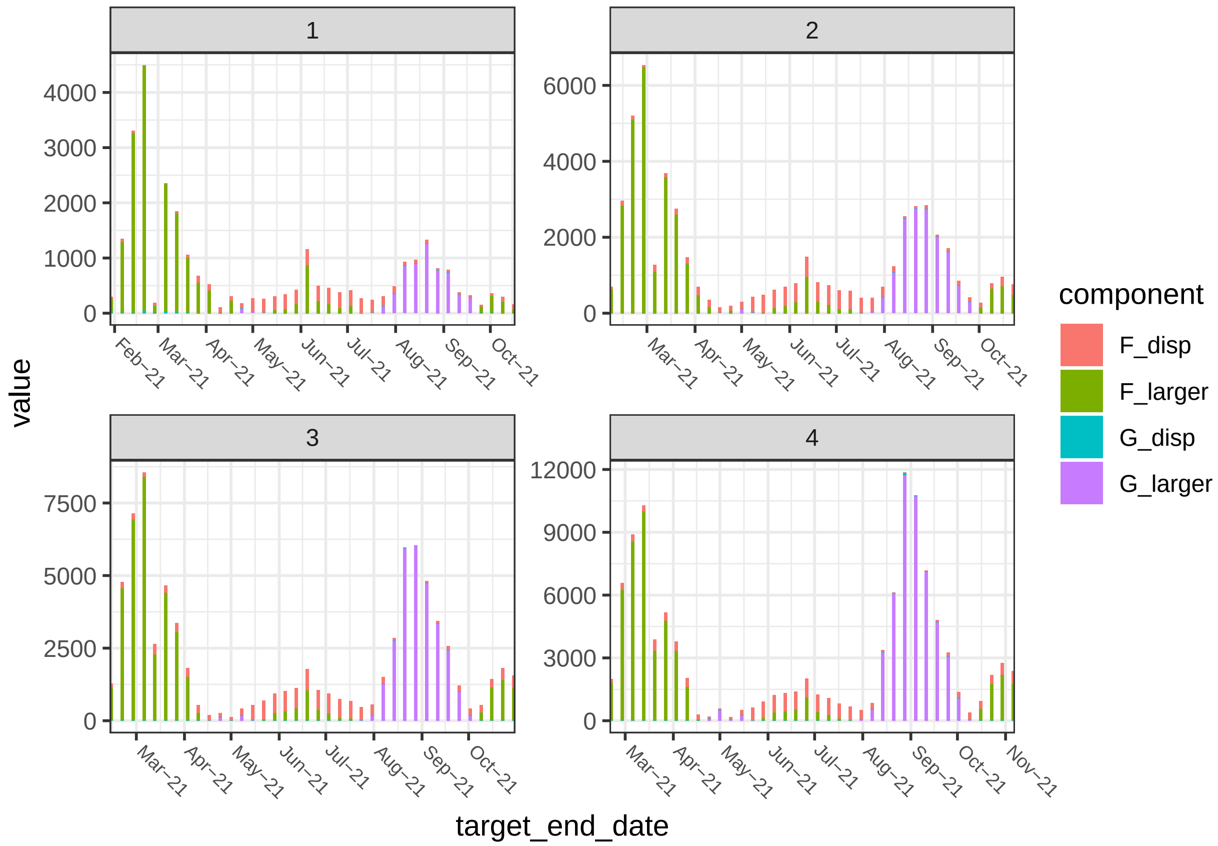
CEID-Walk, RW-ESG, and the baseline are categorized as stats models. Here we plot the distance by date for each location and horizon. We can compare the value vertically (within the same date), but not across the dates since they are pretty much affected by the scale.






Mean Approx. CD of Inc Death Forecasts − TX
1
2
3
4
Mean Approx. CD
pair_type
epi−pair
stats−epi
stats−pair

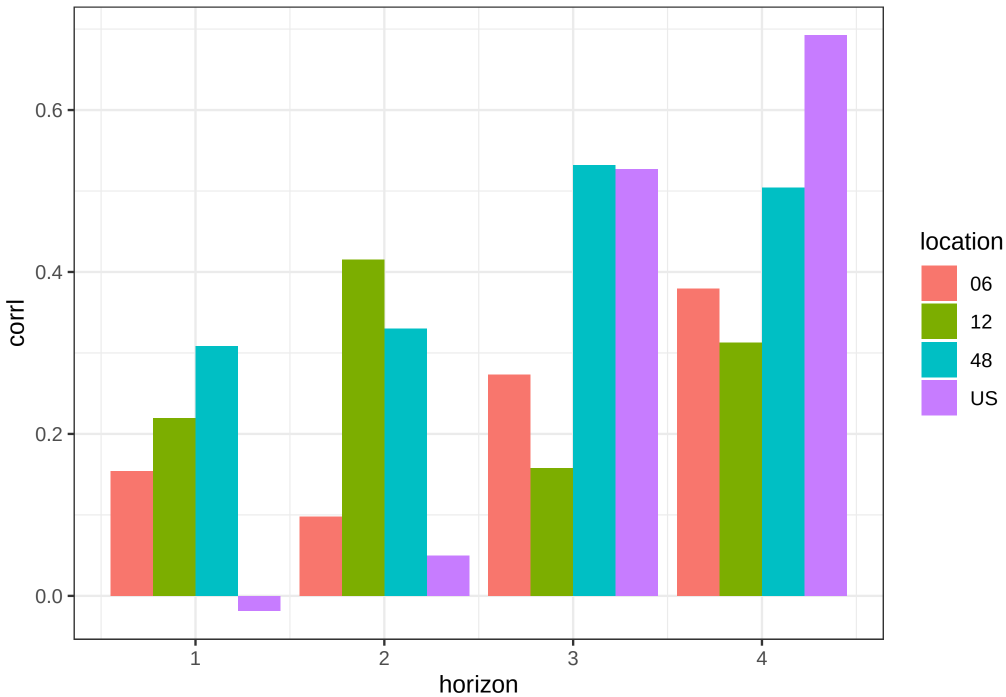

Mean Approx. CD of Inc Death Forecasts − US
1
2
3
4
Mean Approx. CD
pair_type
epi−pair
stats−epi
stats−pair



- We don't see clear differences at any particular time in CA, but we do for other locations
- For FL, stats-pair dissimilarity is high compared to other at times when we observed an increase in death. This might be from RW-ESG since we see that (though not directly) from the box plot - and we have less stats models, so it's more pronounced.
- For US, we see a lot of dissimilarity among stat-epi pairs and between stats pairs at the downward slope after winter wave. At the upward slope in the fall, we see a lot of dissimilarity among epi-pair and stat-epi-pair (does this just imply epi models saying different things?) at 3- and 4-wk ahead horizons (compared to stats pair that seem to agree with each other in that period of time).
- Decomposition for the US between baseline (F) and Karlen-pypm (G), at the downward slope after the peak, baseline model mean being larger than Karlen's mean accounts for most of their dissimilarity. At the beginning of the fall peak, Karlen's mean being larger accounting for most of their dissimilarity. During the flat period, we see differences in their dispersion comes into play.

Variations in distances from relative to baseline to relative wis

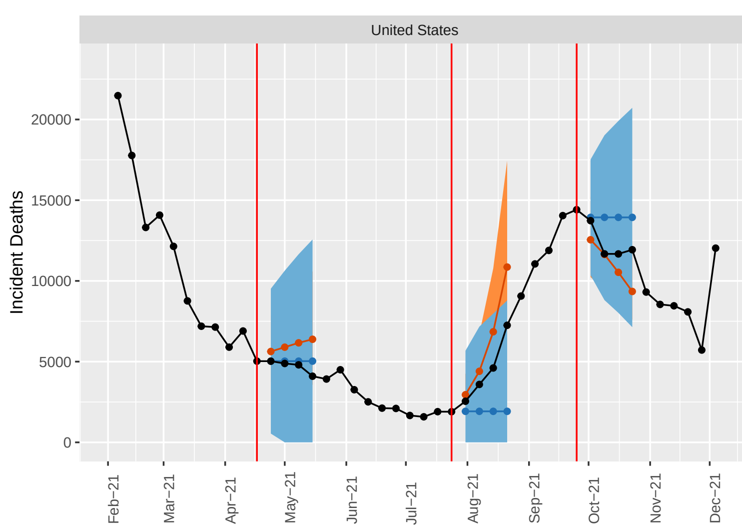
- more disagreement - lower accuracy

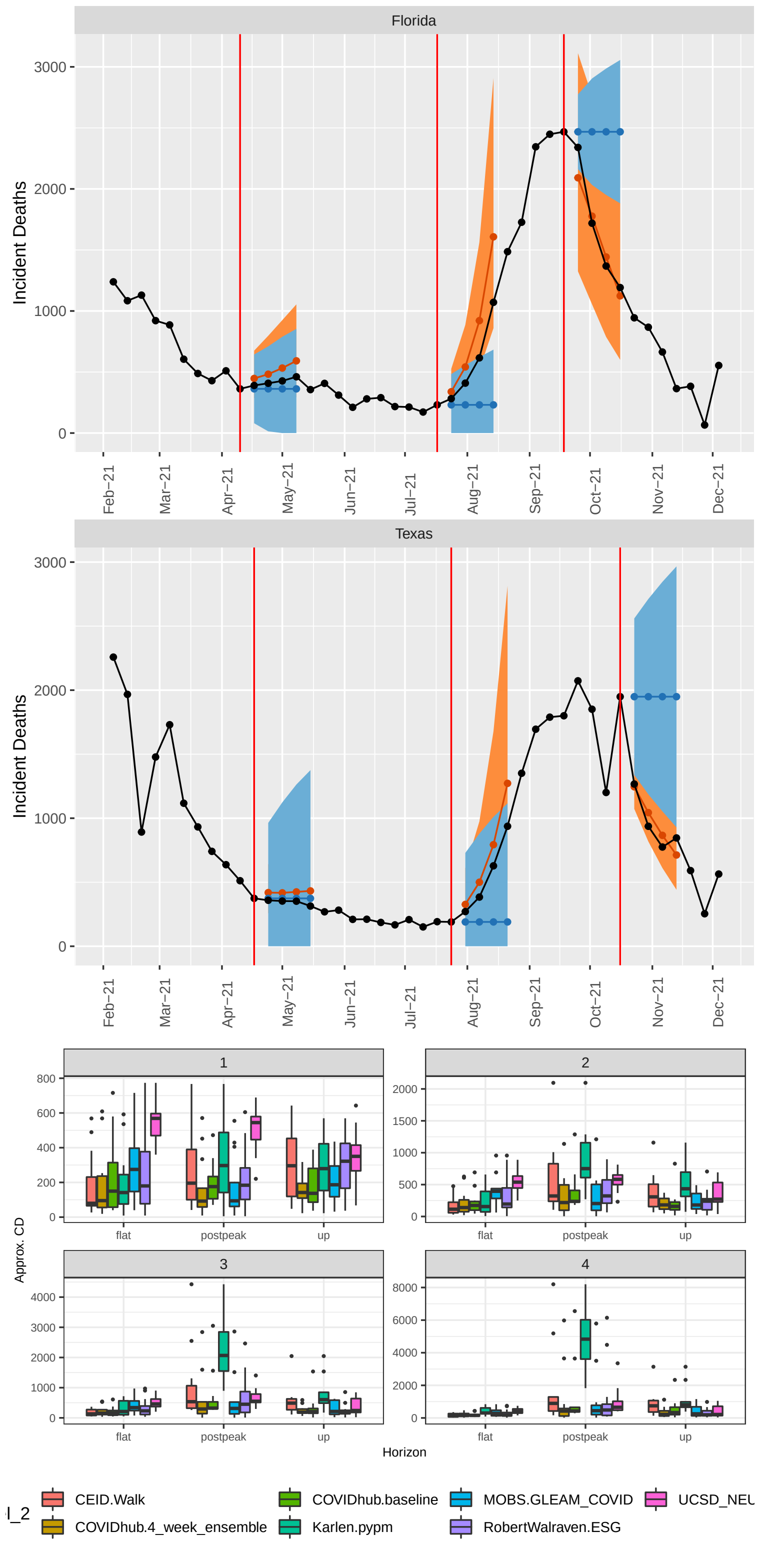


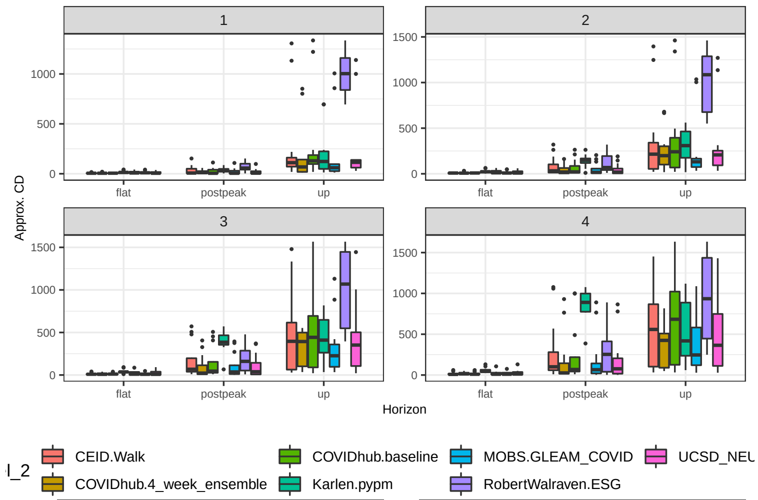
- for each horizon,location, end_date, calculate the mean of each model's cd to other models (except baseline) and divide by mean of baseline's cd to other models and use as a scaled cd for that model. So, we have all models' mean scaled cds at each horizon,location, end_date. Then find standard error of the means for each horizon, location, end date. Then find correlation between SE and wis relative to baseline.
- the more models disagreeing with others relative to baseline's disagreement with others (higher variation in distances), the higher wis are relative to baseline (they are less accurate compared to baseline)...what does this mean?
- disagreement is the varying degree to which model A,model B, ... are dissimilar to others compared to the degree baseline is dissimilar to others.

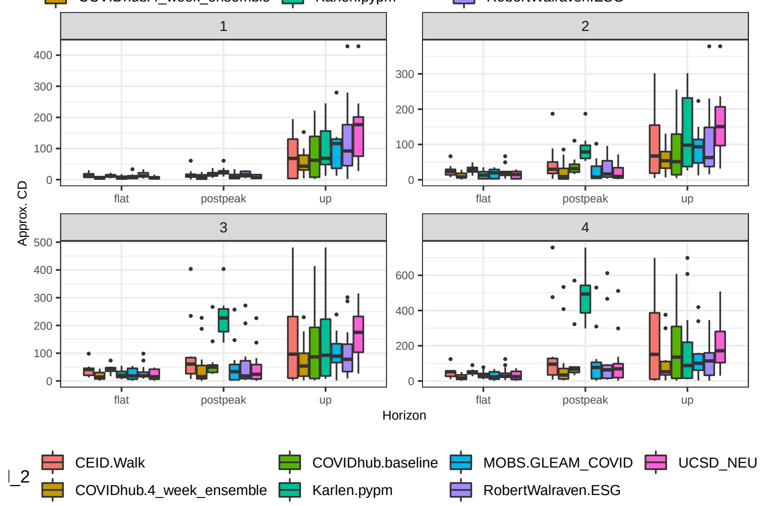
Phase analysis (peak and valleys and flat)

Do it for TX,FL, US since they have discernable peaks. Look at the red line for reference points.

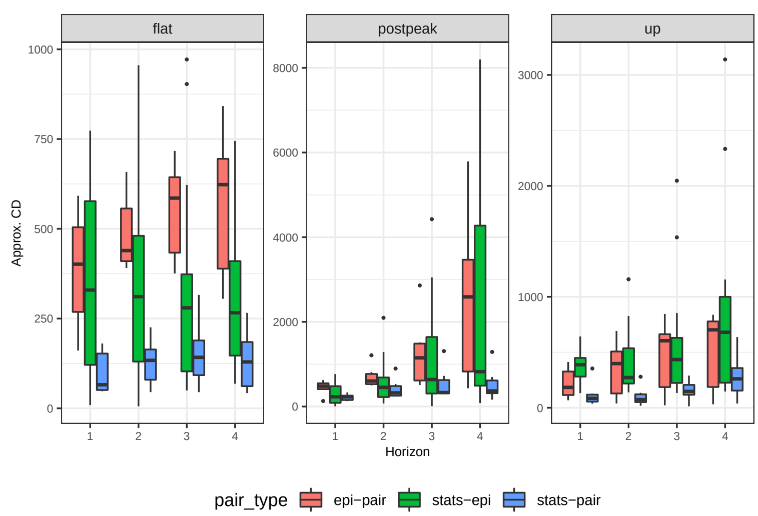


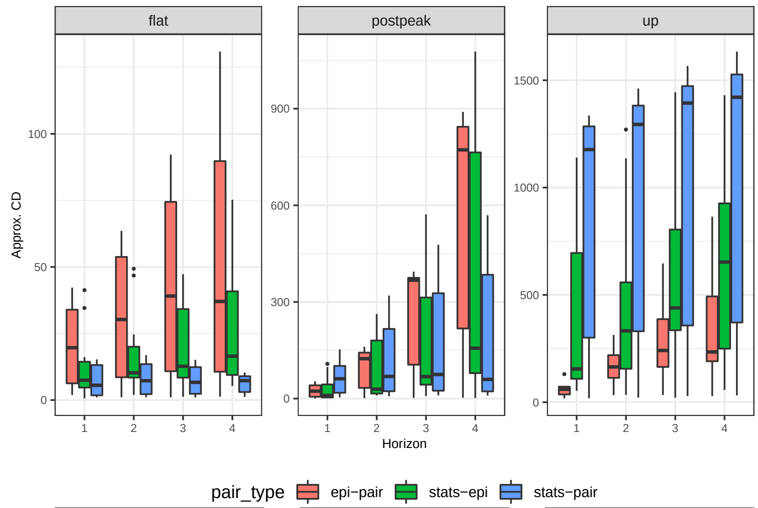

Florida
Incident Deaths
3000
2000
1000
0
Feb−21
Mar−21
Apr−21
May−21
Jun−21
Jul−21
Aug−21
Sep−21
Oct−21
Nov−21
Dec−21
Texas
1
2
3
4
Approx. CD
flat
postpeak
up
Horizon
CEID.Walk
COVIDhub.4_week_ensemble
COVIDhub.baseline
Karlen.pypm
MOBS.GLEAM_COVID
RobertWalraven.ESG

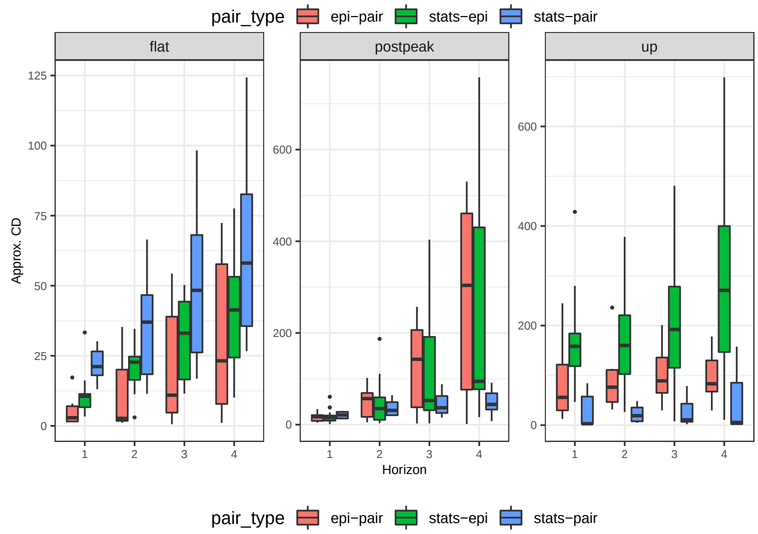




- probably should change the plots' x axis for FL since the up phase stretch the scale.
- The plots with models show each model's average similarity to other models.
- We see differences at each phases in each location - a lot to unpack here later







- In TX, stats-epi pairs' dissimilarity is high when we are in the upward slope phase of the pandemic. Epi models show sign of confusion post peak (downward slope)
- In FL, high variation in similarity among epi models when the death numbers are stable, strange?

Consistancy (Dissimilarity and COVID-19 characteristics)

Consistency of forecast sequences (forecasts from the same model at different horizons made for the same target end date) using the divergence index (based on cramer distance) proposed by Richarson et al. (2020). The difference between two forecasts made on consecutive weeks and valid for the same time is

$$D_{enddate=t,h} = cd(F_{enddate=t,h+1}, F_{enddate=t,h}), h = 1, 2, 3$$

Divergence index is

$$DI_{enddate=t} = \frac{1}{3}(\sum_{h=1}^{3} D_{enddate=t,h} - cd(F_{enddate=t,3}, F_{enddate=t,1}))$$