COVID-19 Forecast Similarity Analysis

Johannes Bracher, Evan Ray, Nick Reich, Nutcha Wattanachit

10/25/2021

Goals

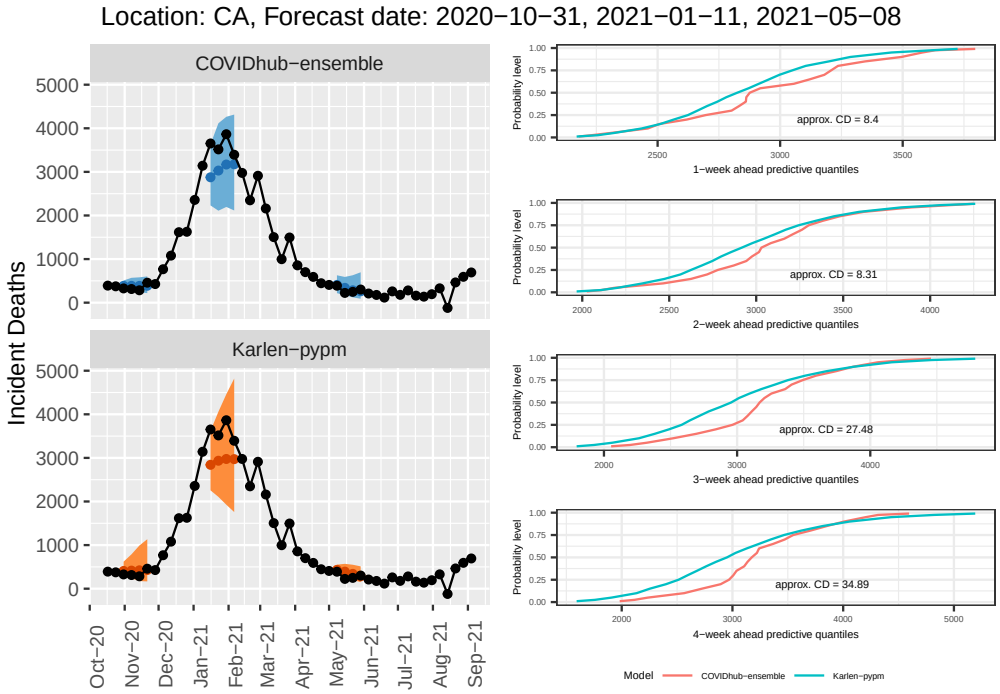
- Check model types and their dissimilarity on average (heatmaps and boxplot)
- Check the above for inflection points and stable periods (can we use differences to predict inflection point?), rather than inflections, maybe the begining and the end of a wave…or maybe think of this as onsets, peaks, end of waves??..
- Decomposition by type analysis
- Test for difference in group mean dissimilarity (small sample,repeated measures,non-normal, maybe unequal group samples for cases) - maybe if we don't aggregate and treat these like repeated measures in a mixture structure to increase n
- Note: ensemble excluded.
- Test if diversity in an ensemble is associated with performance (simulations).

Forecast inclusion criteria

- Models: All models with complete submissions for the following criteria (applied to death and case forecasts separately)
- Targets: 1-4 wk ahead inc death and inc case
- Target end dates: Oct 19th, 2020 - Sep 18th,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

1-4 Week Ahead Incident Death Forecasts

There are 9 models that fulfilled the criteria for the 5 locations with highest cumulative deaths. Below is an example of approximated Cramér distances between COVIDhub-ensemble and Karlen-pypm for 1-4 week ahead forecasts of incident deaths in CA in the week of 01/11/2021.



Model	Mechanistic
COVIDhub-baseline	FALSE
COVIDhub-ensemble	FALSE
CU-select	TRUE
Karlen-pypm	TRUE
LANL-GrowthRate	FALSE
MOBS-GLEAM_COVID	TRUE
SteveMcConnell-CovidComplete	FALSE
UA-EpiCovDA	TRUE
UCSD_NEU-DeepGLEAM	FALSE

Mean approximated pairwise distances across all forecast weeks

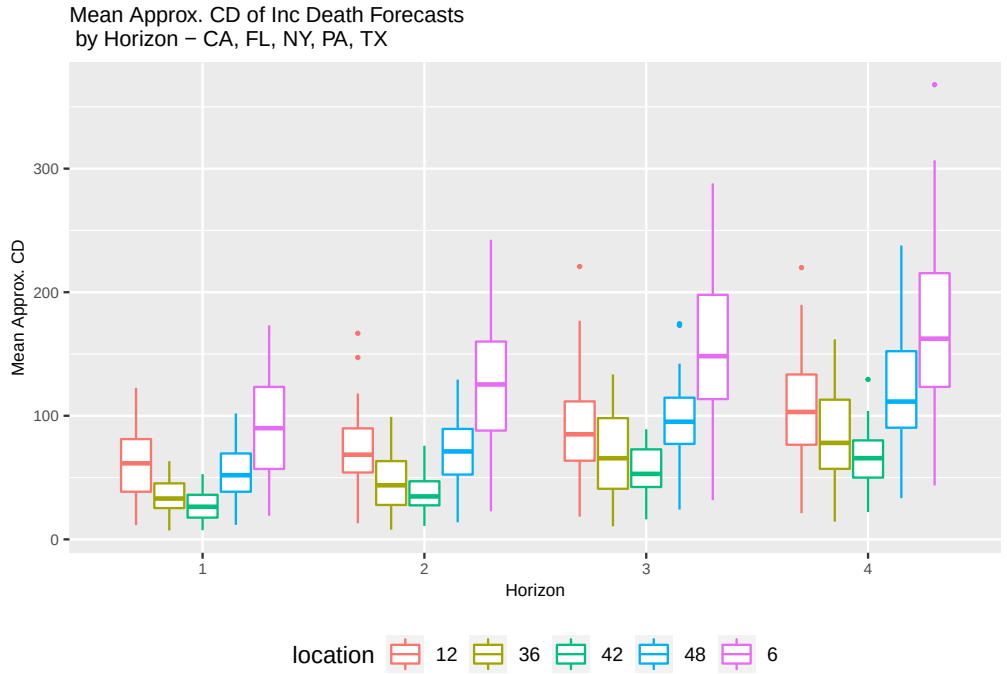
The $x-$axis is arranged based in an ascending order of the model's approximate pairwise distance from the COVIDhub-ensemble. So, the first model is the model that is most dissimilar (on average across all horizons) to the ensemble in this time frame. The observations below are based on the plots. Note the scale of approx. CD is much smaller for the locations with low cumulative deaths (despite the same color scale).

For locations with high cumulative deaths:

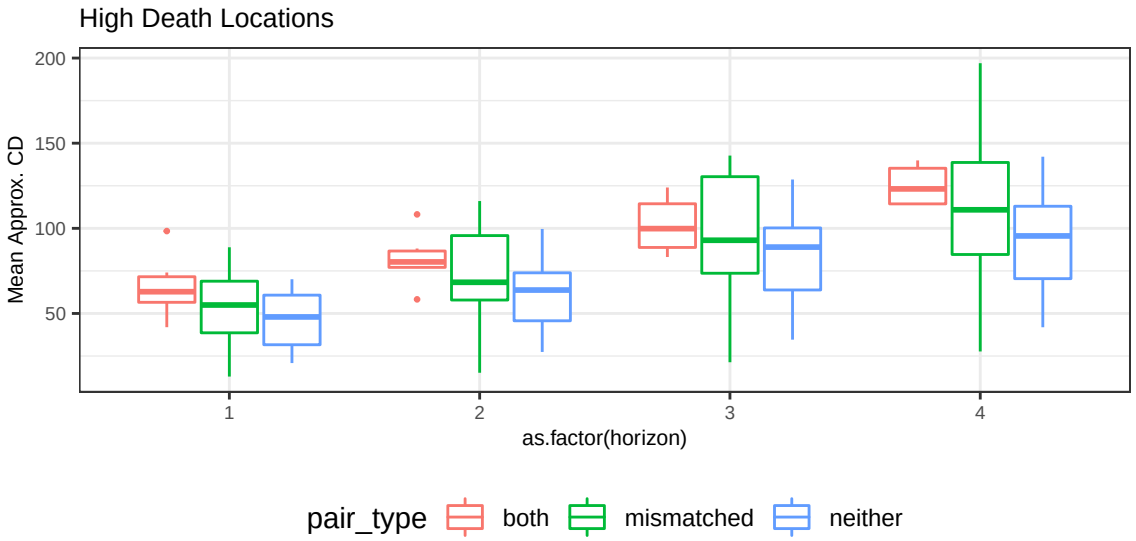
- The mean approx CD between pairs of forecasts are higher for further forecast horizons
- CU-select is, on average across all horizons, most dissimilar to the ensemble
- CU-select seems to be most dissimillar to other models for 1-2 wk ahead horizons, while UA-EpiCovDA is most dissimilar to other models for 2-4 wk ahead horizons. For locations with low cumulative deaths:
- We see only little bit of the trend in the first bullet point
- For all horizons, UCSD_NEU-DeepGLEAM is noticeably most dissimilar from other models, UA-EpiCovDA is perhaps the second most dissimilar.

From the boxplots by location:

- Less outliers for locations with high cumulative deaths
- For locations with high cumulative deaths, while the range of approx. CDs for all unique model pairs might be different, the relationship between locations are about the same across all horizons. So, averaging across locations to reduce the granularity makes sense. Though we might not be able to say the same thing for locations with low cumulative deaths, the differences are probably too small for us to care.



A mechanistic model type and similarity



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

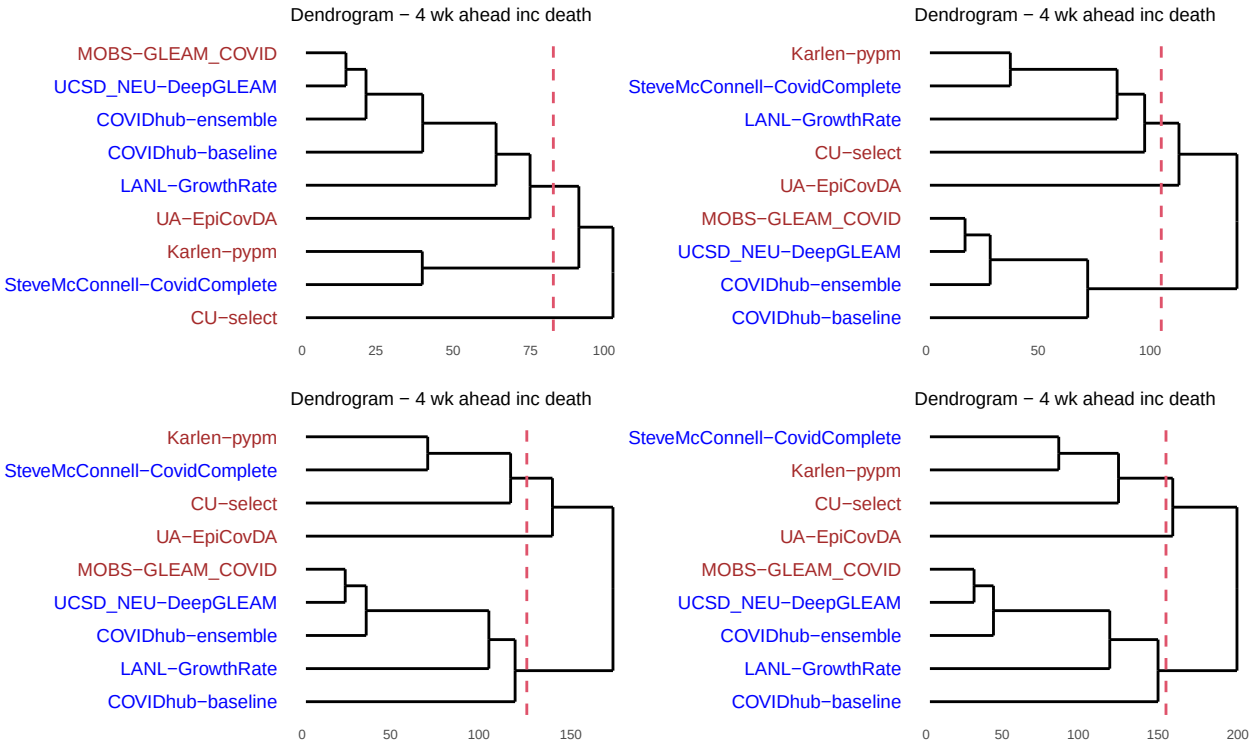
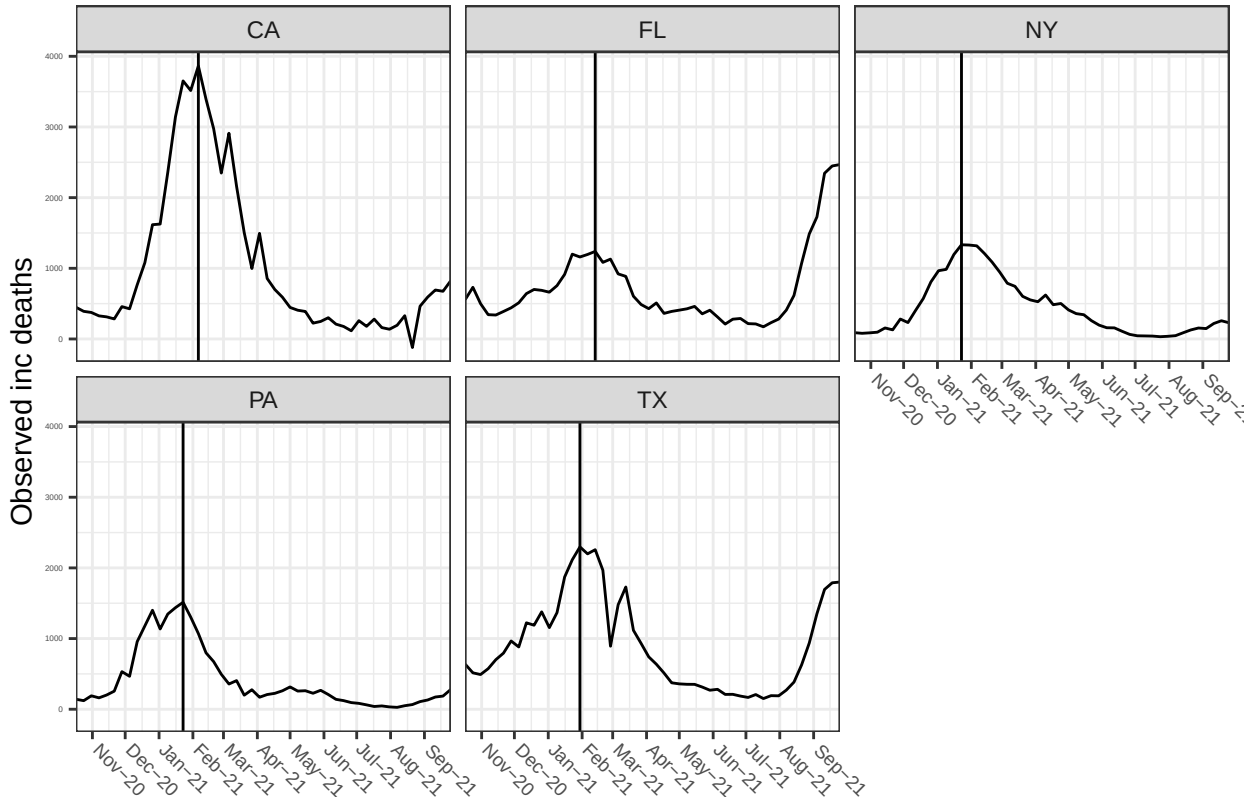
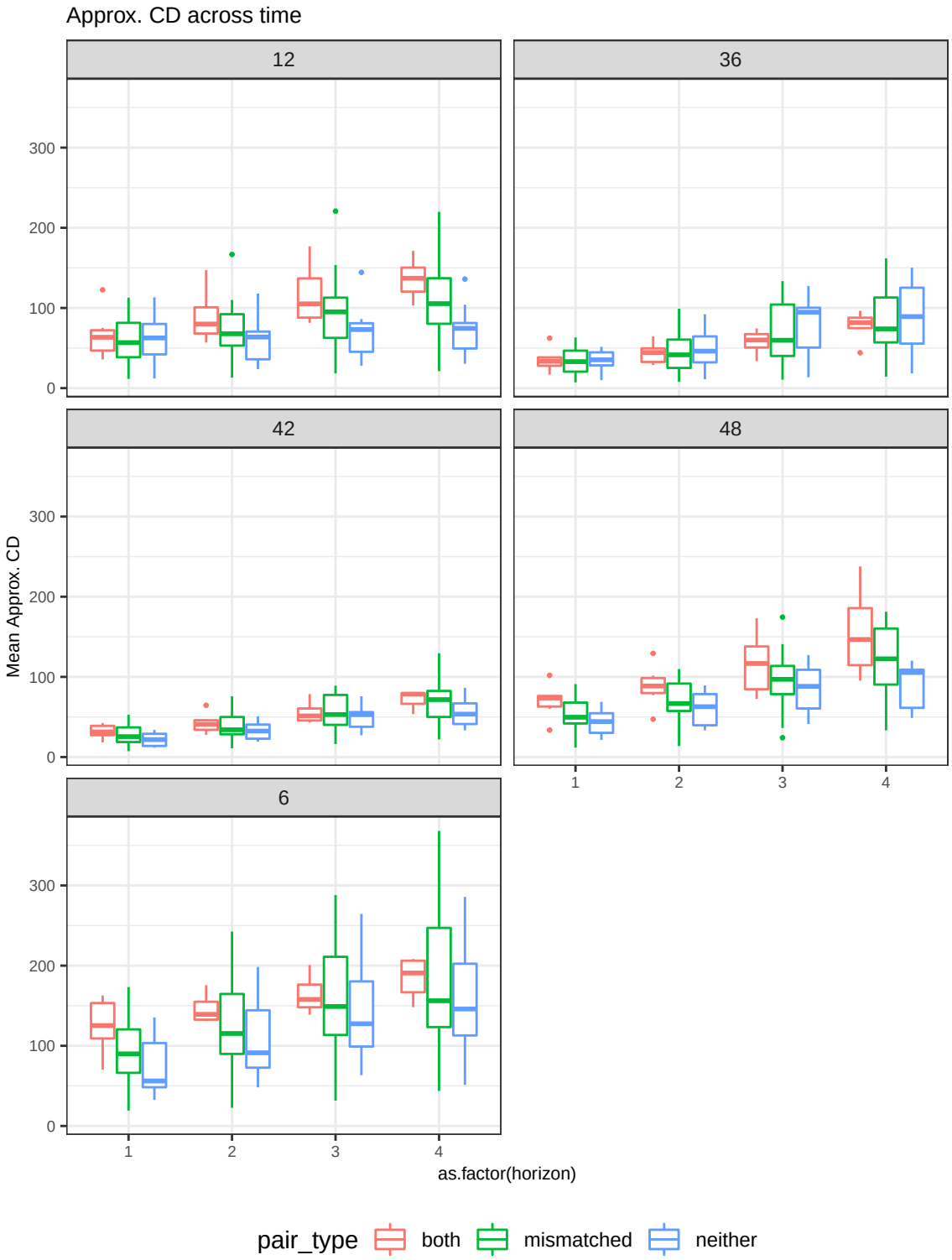
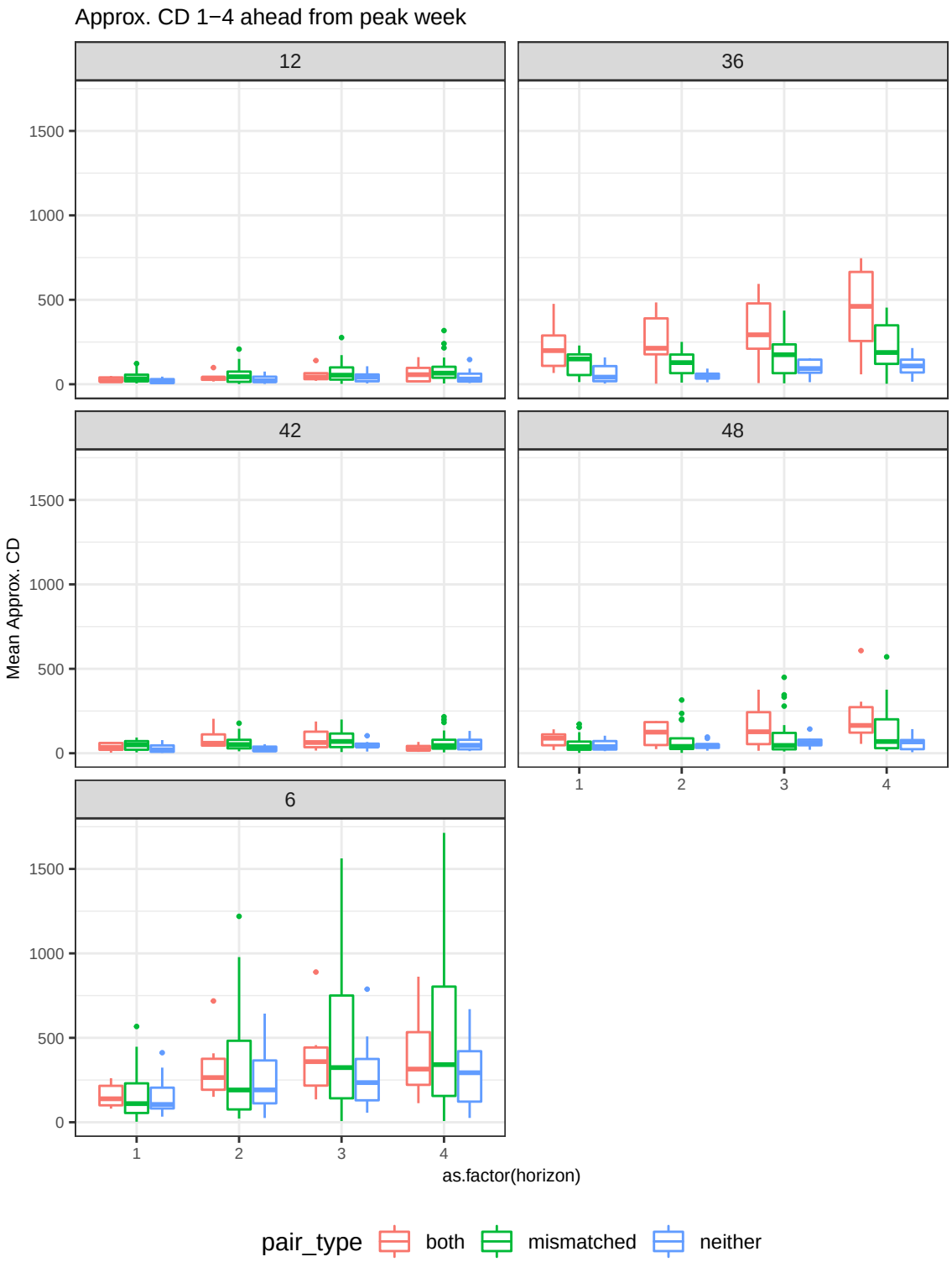


Figure 1: High Case Count Locations

Boxplots and Hierarchical clustering based on mean approx. CD overll vs after peak week




Approx. CD across time
12
36
42
48
6
Mean Approx. CD
0
100
200
300
1
2
3
4
as.factor(horizon)
pair_type
both
mismatched
neither


Approx. CD 1−4 ahead from peak week
12
36
42
48
6
Mean Approx. CD
0
500
1000
1500
1
2
3
4
as.factor(horizon)
pair_type
both
mismatched
neither

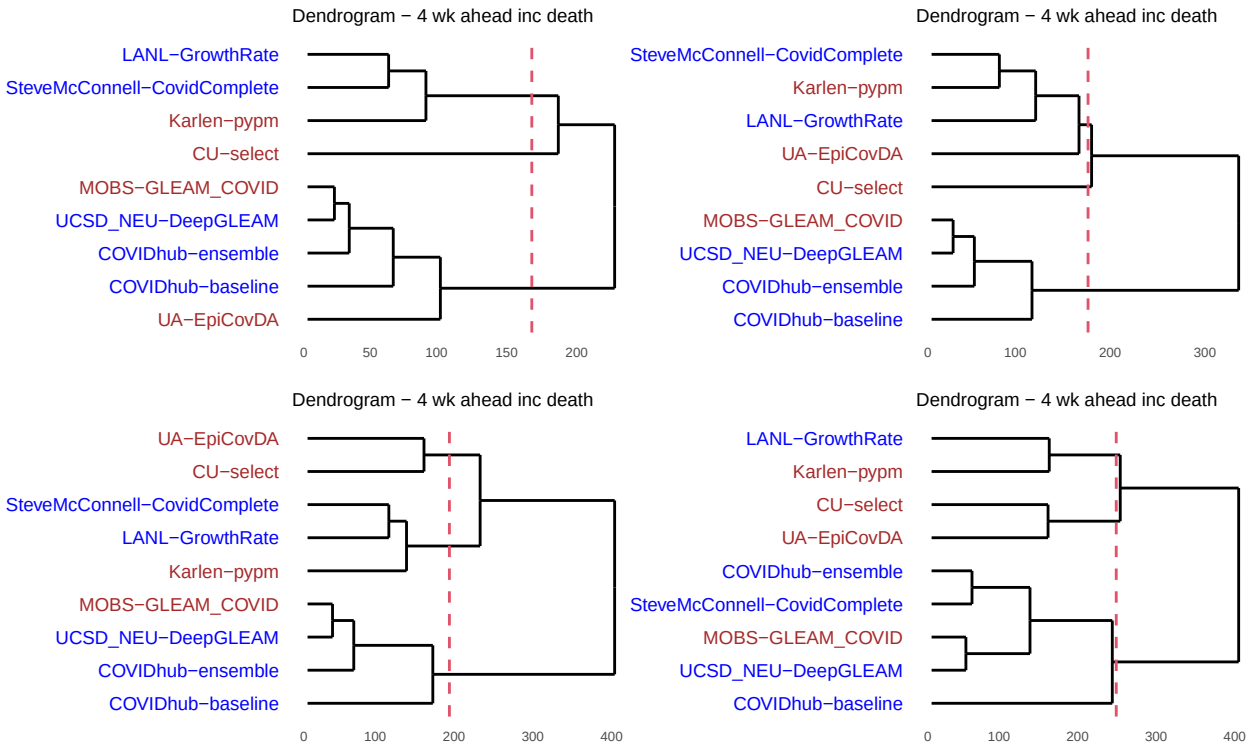


Figure 2: Location 12 - across time

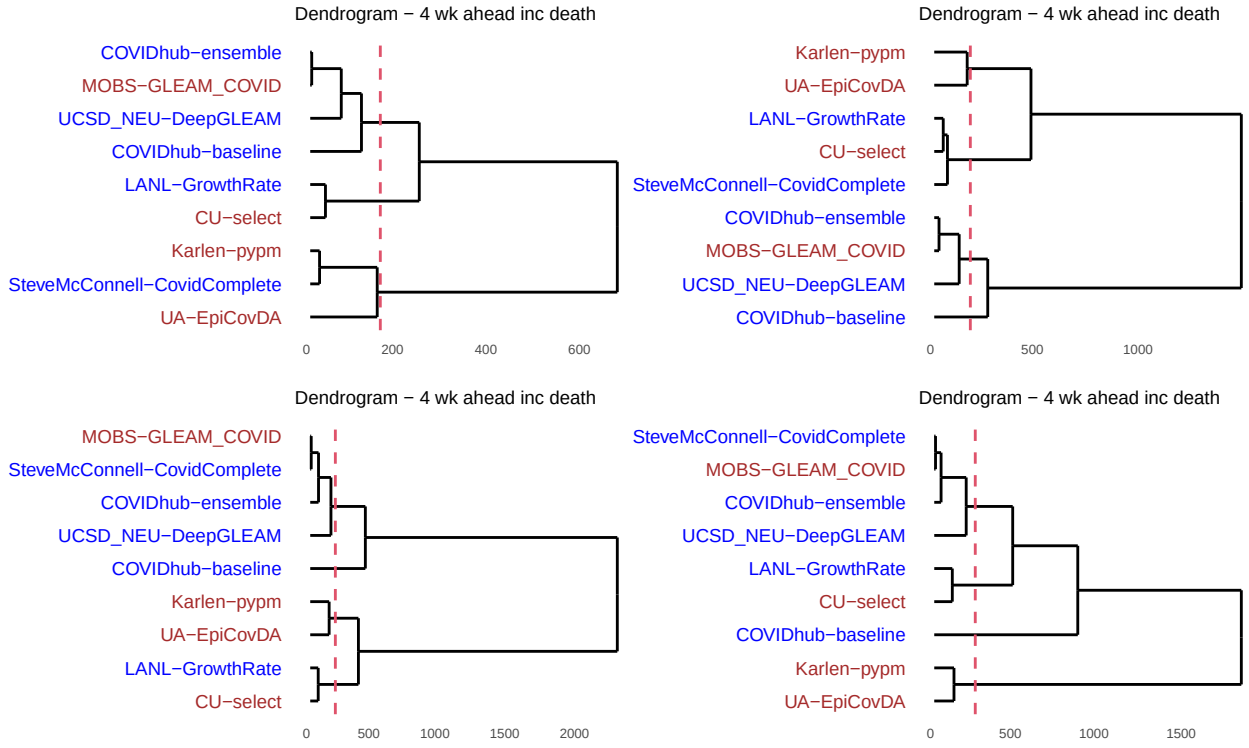
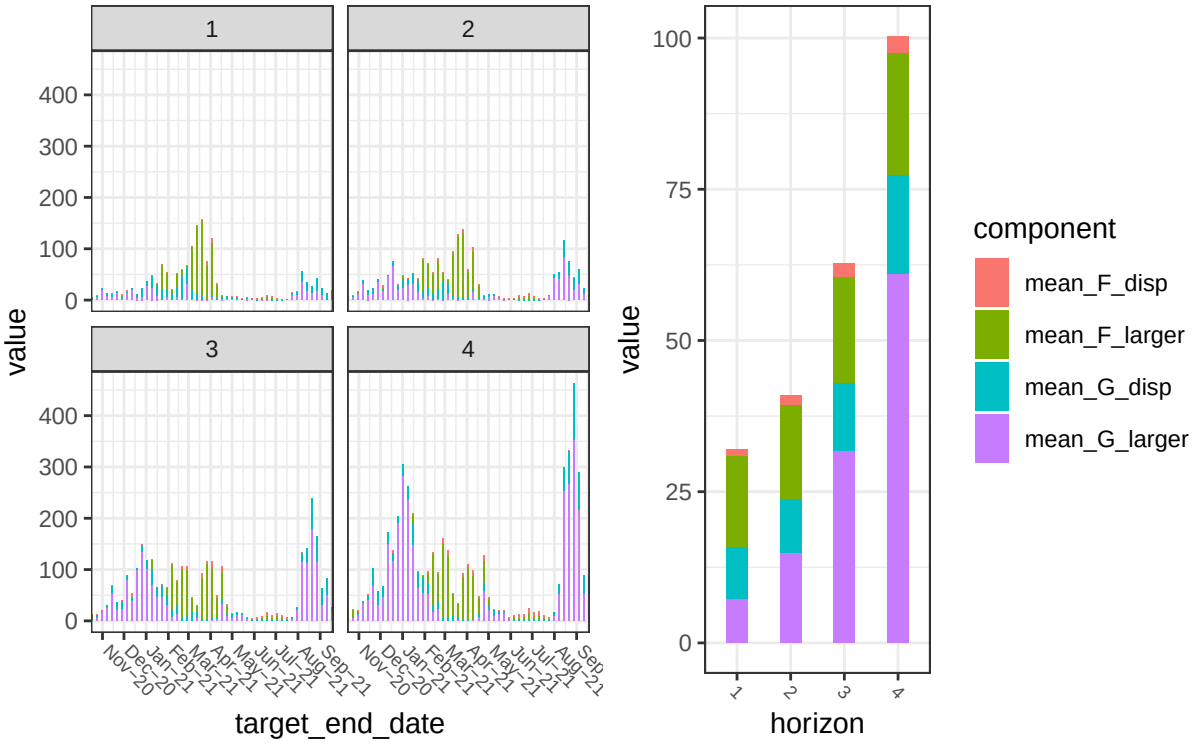


Figure 3: Location 12 - 1-4 week from peak week

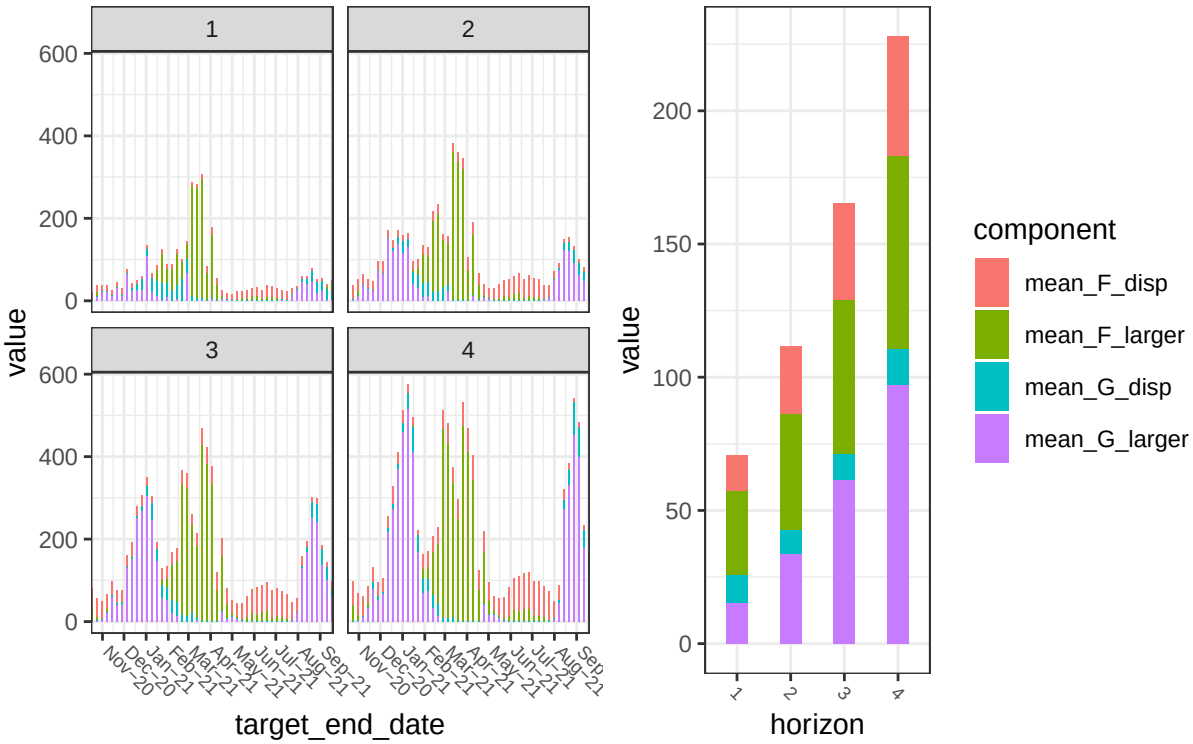
Decomposition

This is a trial run of applying decomposition to the actual forecasts. For now, we show the decomposition of the approx. CD between the COVIDhub-ensemble (Model F) and Karlen-pypm (Model G).

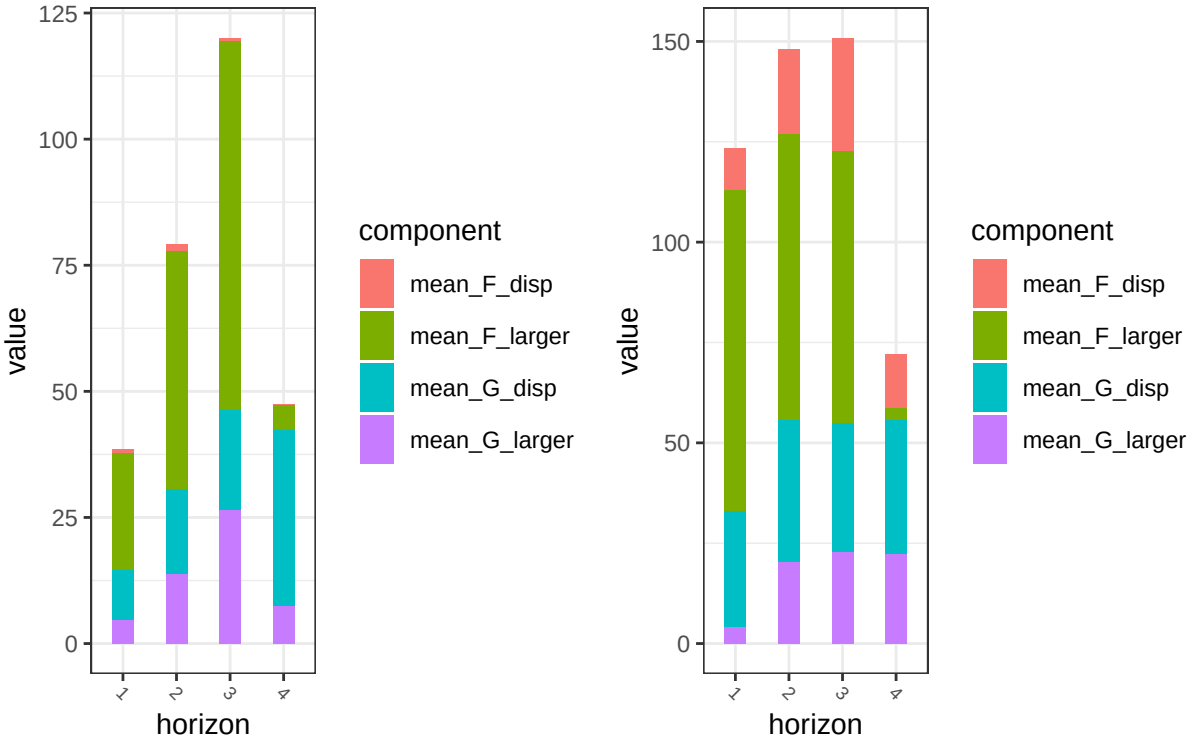
- Both dispersion components seem to play little role in dissimilarity
- Comparing between forecasts for locations with high deaths vs ones with low deaths, the proportions of decomposition components look about the same, with the upward shift of G relative to F accounting for most of the dissimilarity on average
- Around Feb 2021 until the end of April 2021, we see the upward shift of F relative to G accounting for most of the dissimilarity during that period. This coincides with the downward slope of observed inc deaths after the peak. More inspection needed to discern more information.



Decomposition of approx. CD between COVIDhub−ensemble and Karlen−pypm − CA, FL, NY, PA, TX



Decomposition of approx. CD between COVIDhub−baseline and Karlen−pypm − CA, FL, NY, PA, TX



Decomposition of approx. CD between COVIDhub−ensemble and Karlen−pypm (left) and COVIDhub−baseline and Karlen−pypm (right) − CA, FL, NY, PA, TX

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)