

Forecasting Model Similarity

Johannes Bracher, Evan Ray, Nick Reich, Nutch Wattanachit, Li Shandross

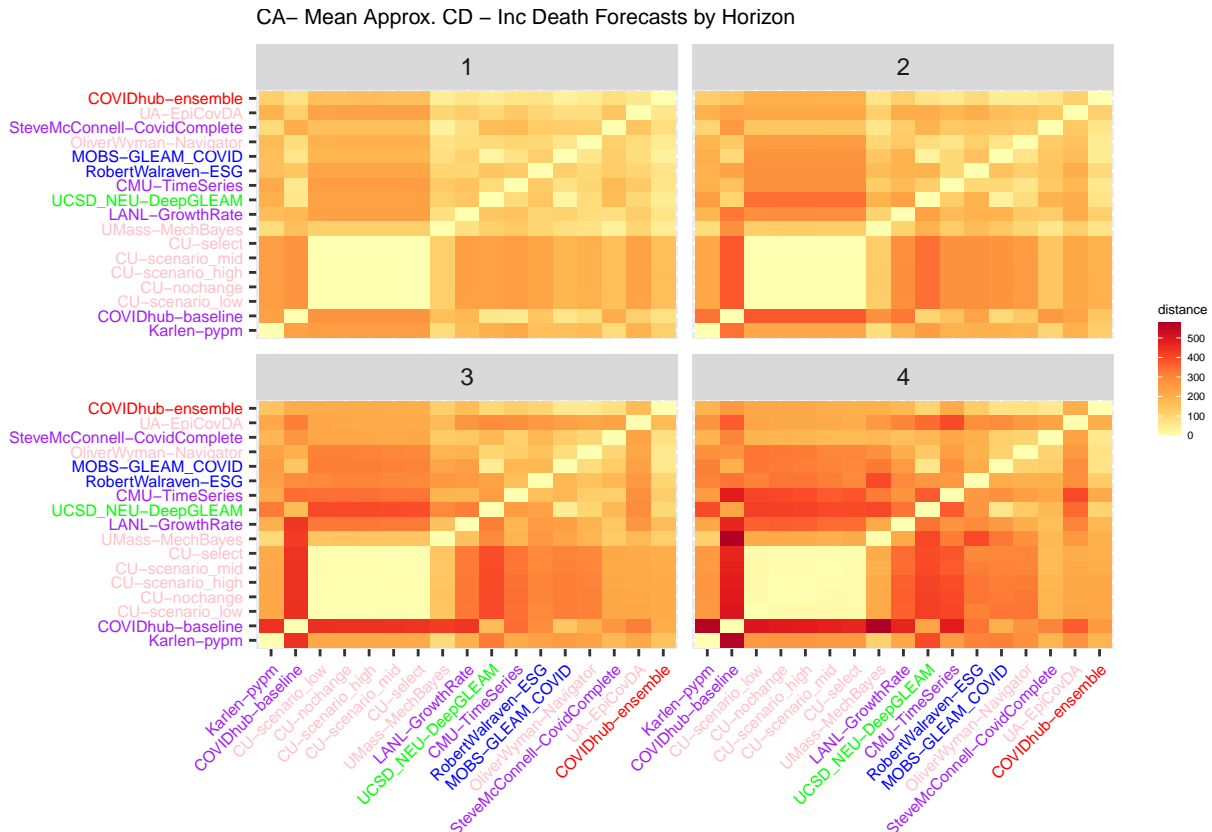
06/21/2021

COVID-19 Forecasting Model Similarity Analysis for 1-4 Week Ahead Incident Death

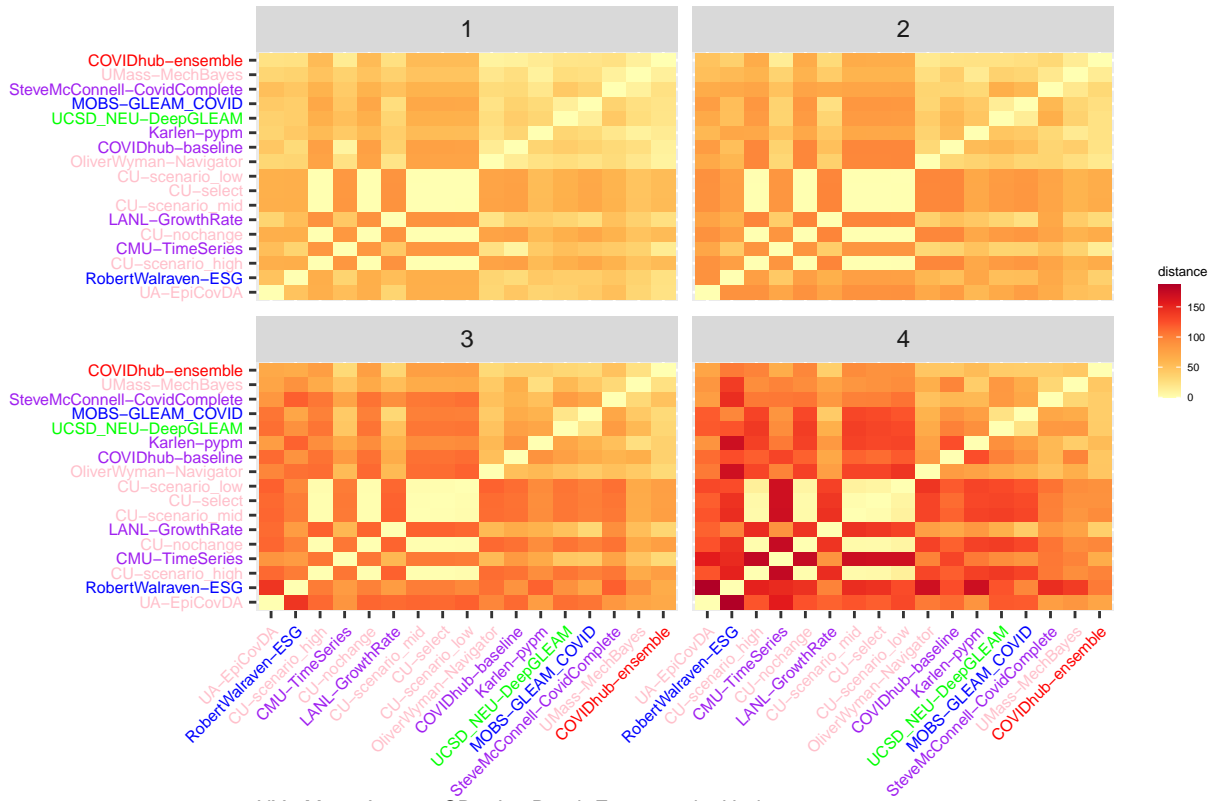
5 locations with the highest number of COVID-19 deaths by the end of February 2021

The pairwise approximated Cramer's distances are calculated for the models that have complete submissions for all target, all 5 locations with the highest number of COVID-19 deaths by the end of February 2021, all probability levels, from the target end date of October, 17th 2020 to May 29th, 2021 .

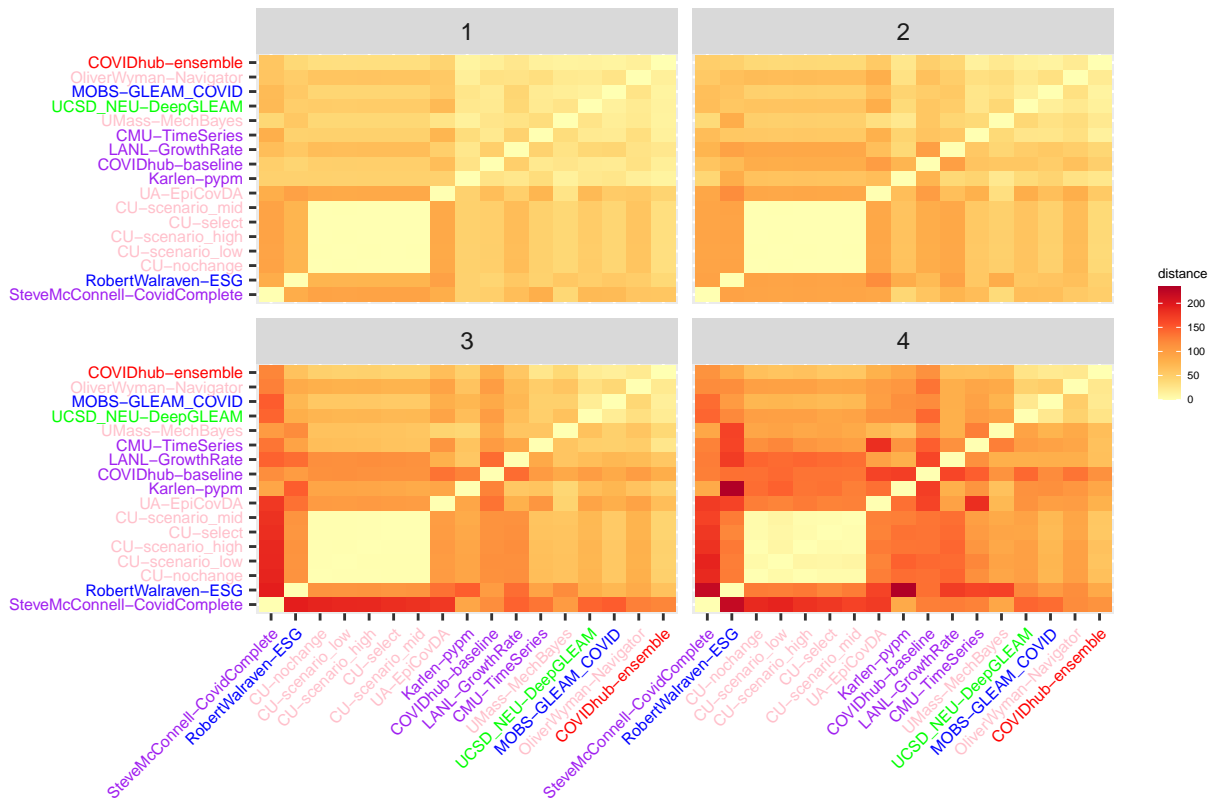
We can visualize the mean approximated pairwise distances across all time points in a heat map shown below. The distance from the model to itself is zero. The x -axis is arranged based in an ascending order of the model's approximate pairwise distance from the COVIDhub-ensemble. So, the first model is the model that is most dissimilar (on average) to the ensemble in this time frame.



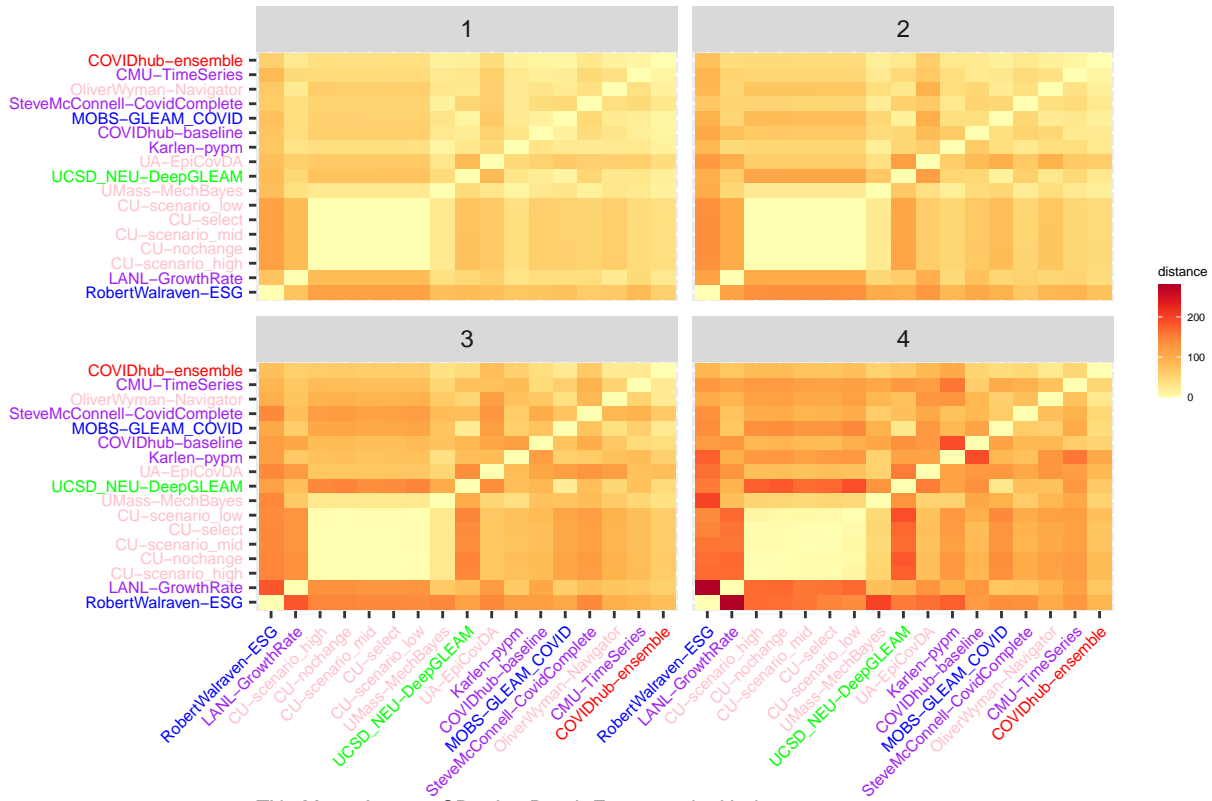
FL– Mean Approx. CD – Inc Death Forecasts by Horizon



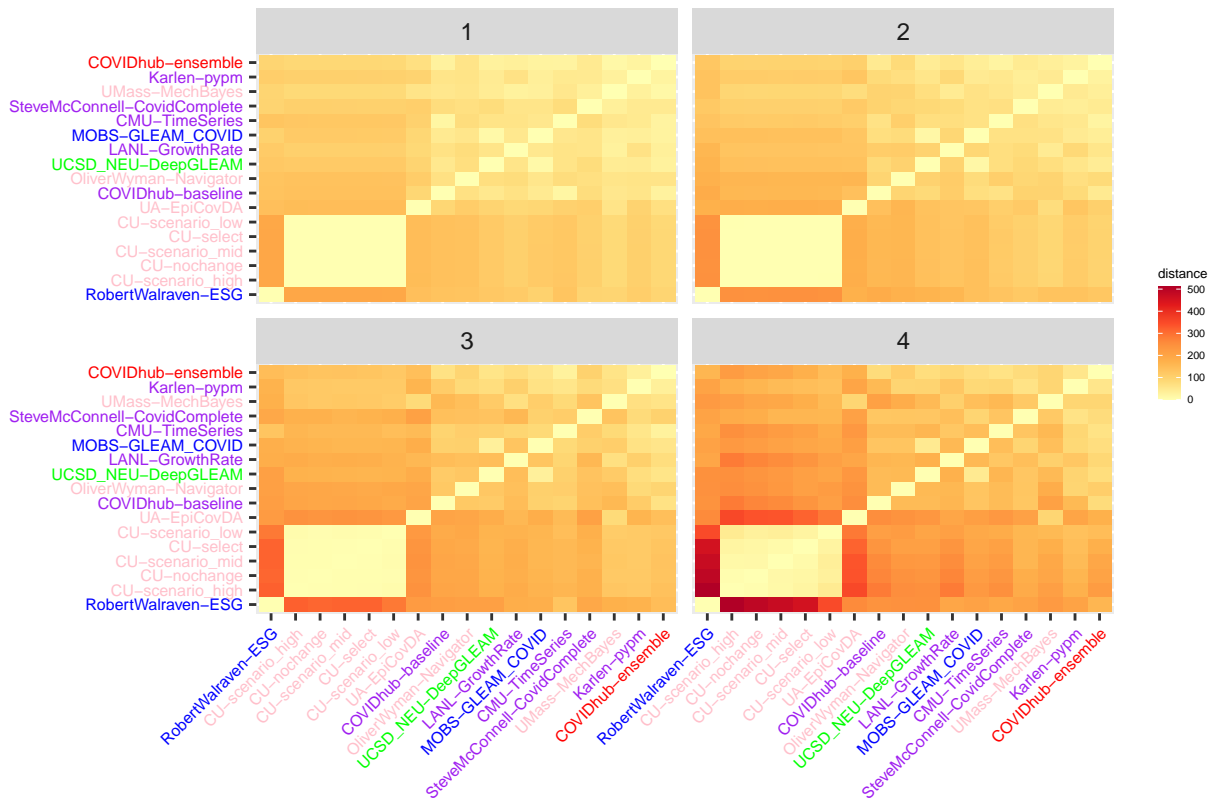
NY– Mean Approx. CD – Inc Death Forecasts by Horizon



PA– Mean Approx. CD – Inc Death Forecasts by Horizon

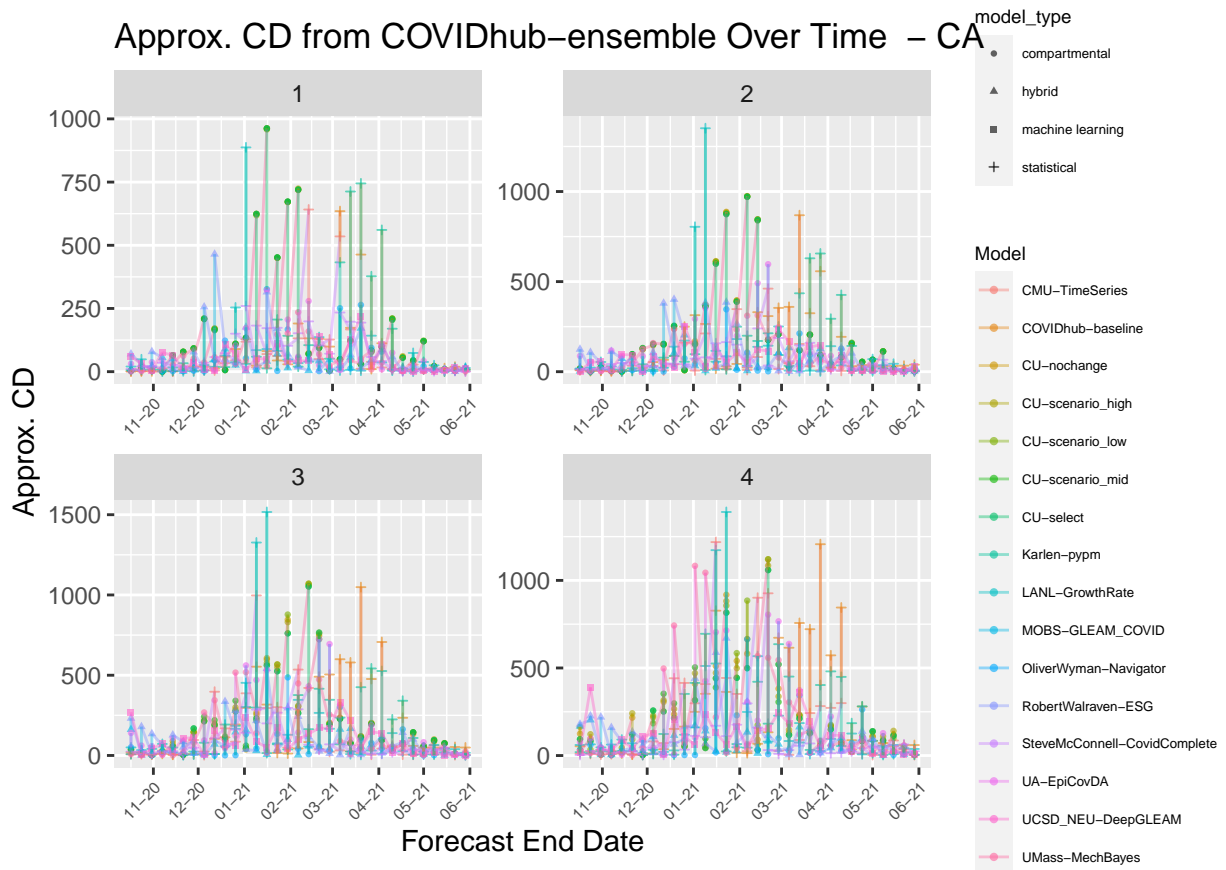


TX– Mean Approx. CD – Inc Death Forecasts by Horizon

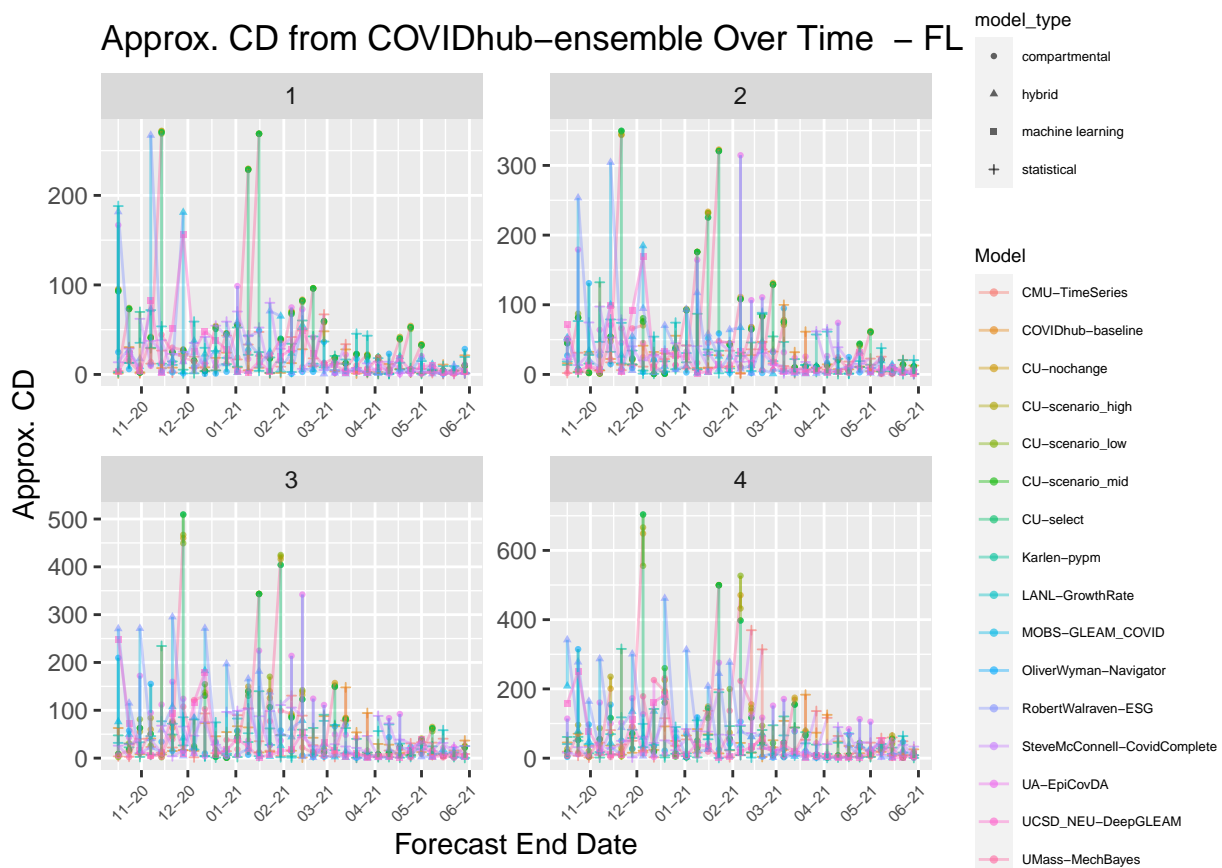


We can also look at the approximated pairwise distances to see how the models become more similar or

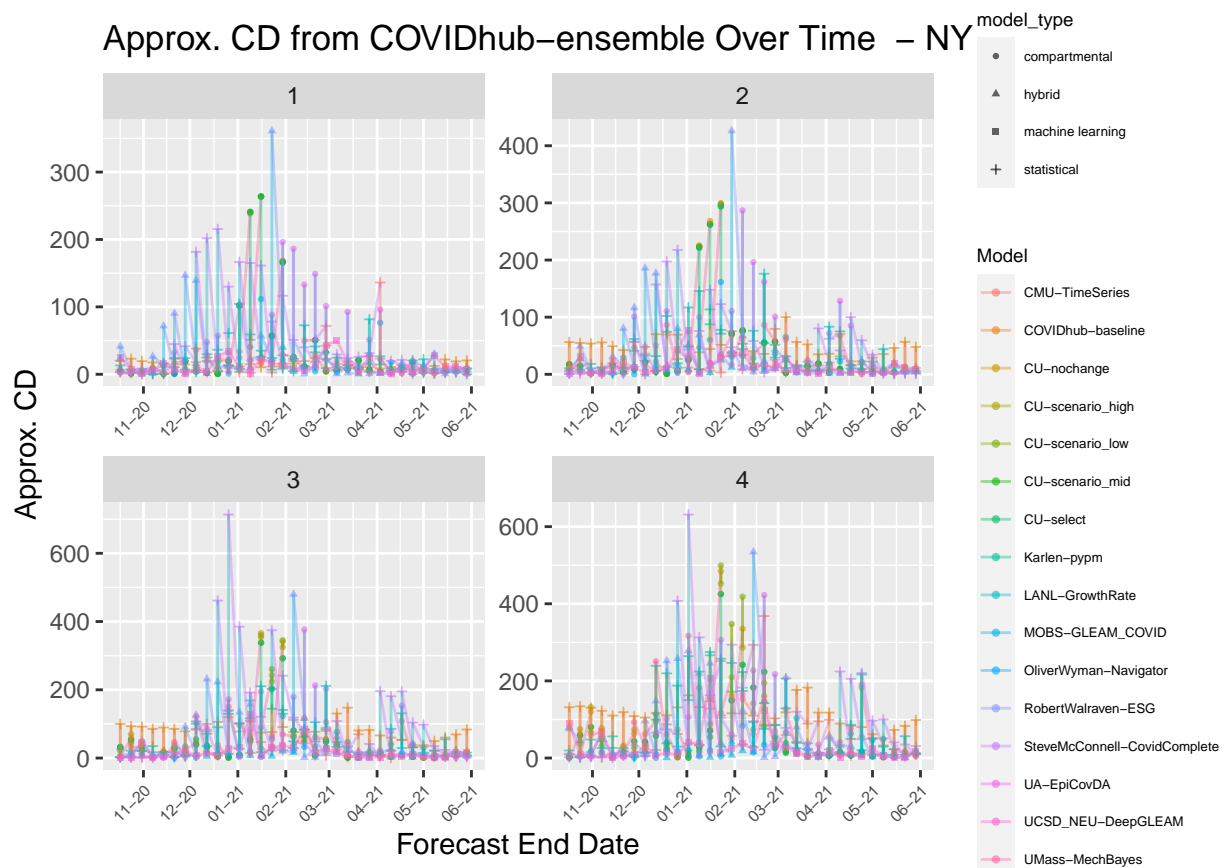
dissimilar over time.



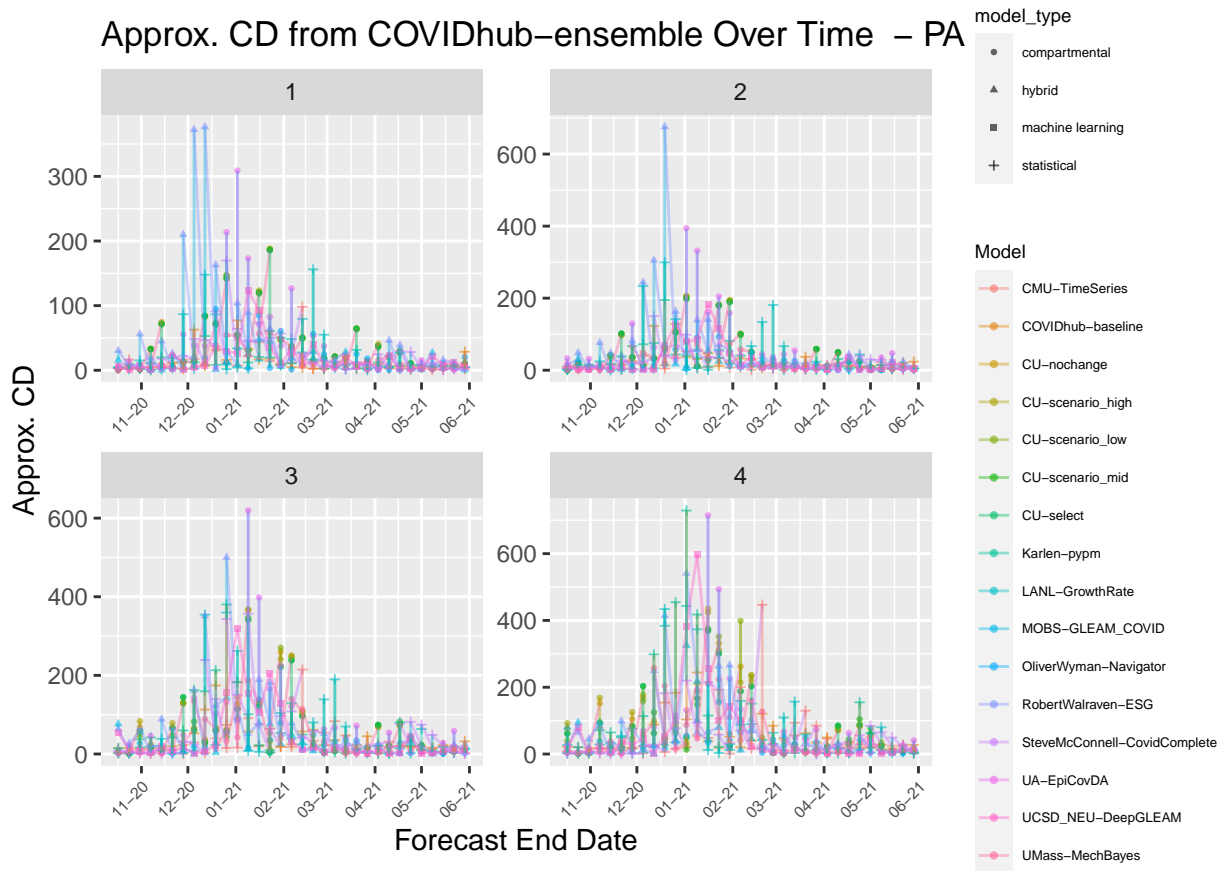
Approx. CD from COVIDhub-ensemble Over Time – FL

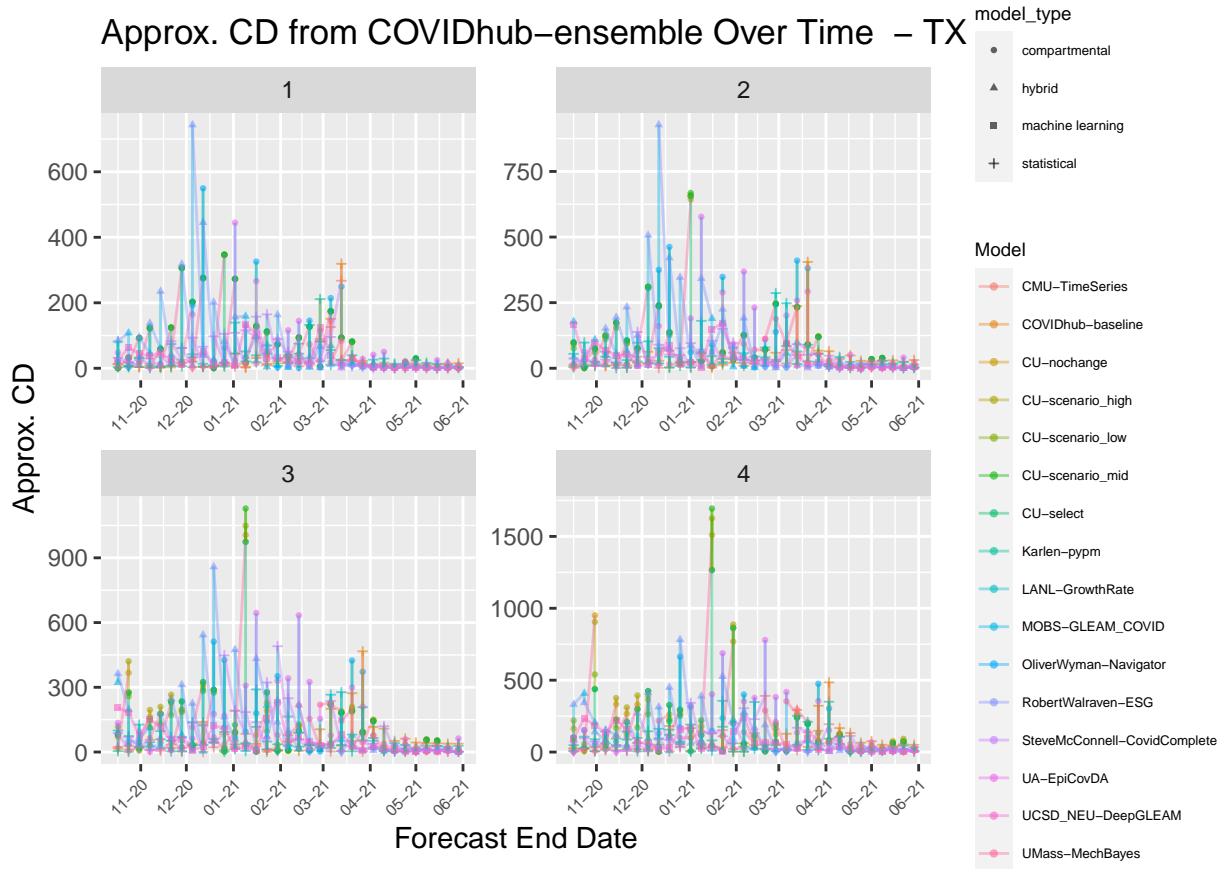


Approx. CD from COVIDhub–ensemble Over Time – NY



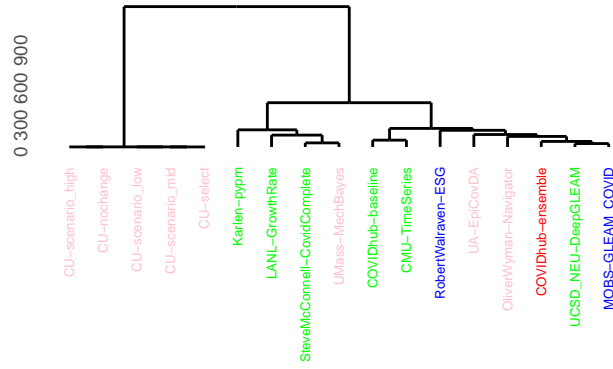
Approx. CD from COVIDhub-ensemble Over Time – PA



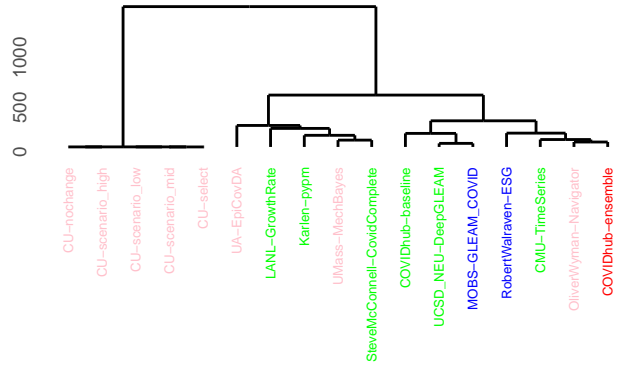


We can cluster the distances using hierarchical clustering. Different linkages will result in different clusters, we probably should investigate more later.

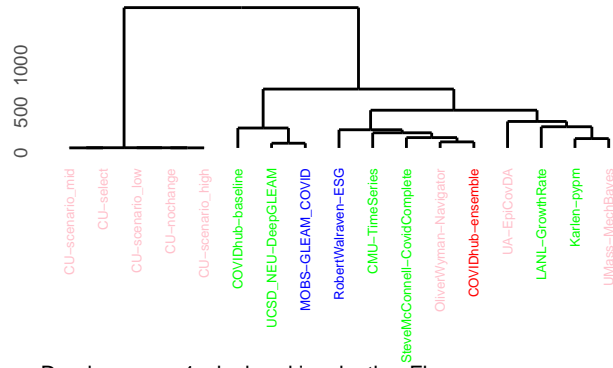
Dendrogram – 1 wk ahead inc death – CA



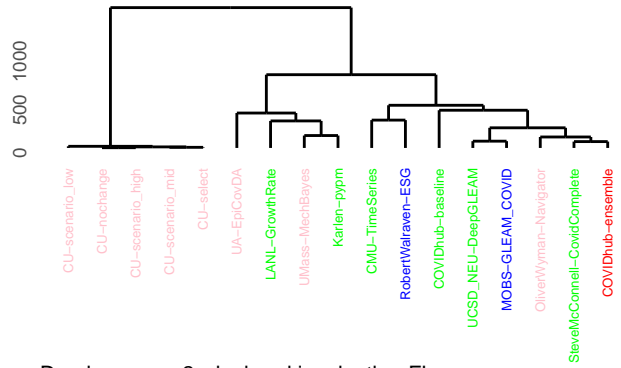
Dendrogram – 2 wk ahead inc death – CA



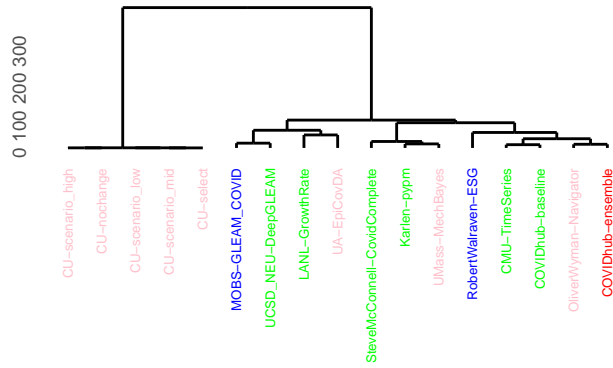
Dendrogram – 3 wk ahead inc death – CA



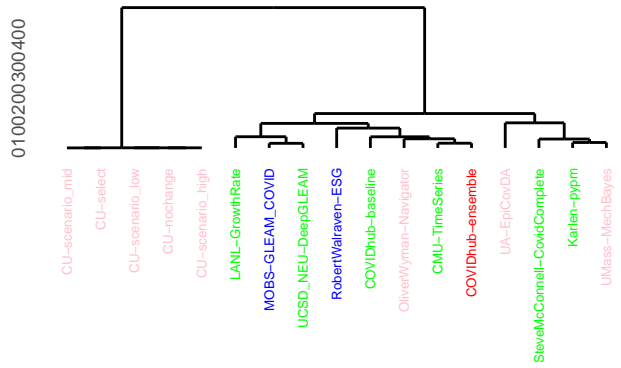
Dendrogram – 4 wk ahead inc death – CA



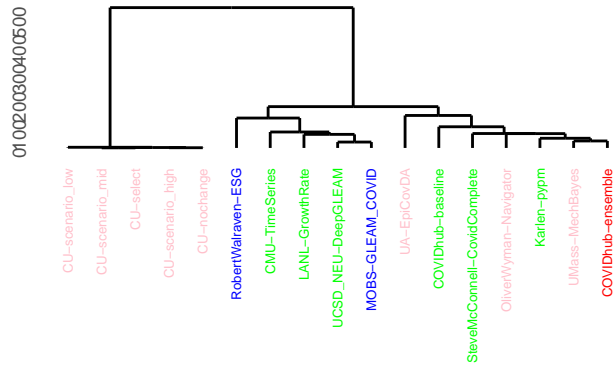
Dendrogram – 1 wk ahead inc death – FL



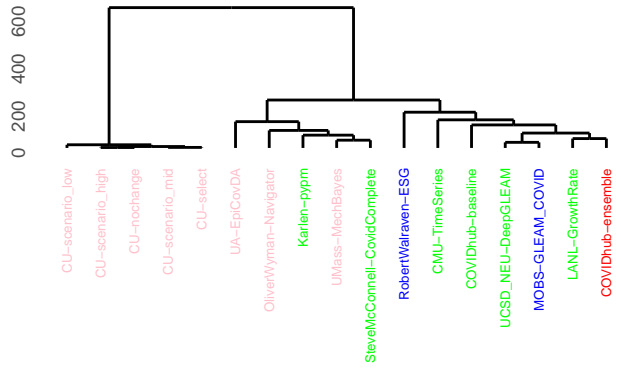
Dendrogram – 2 wk ahead inc death – FL



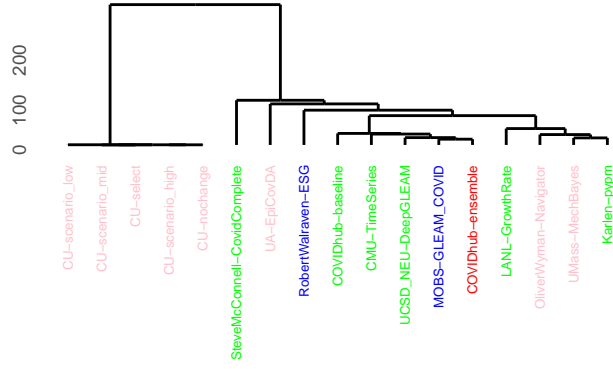
Dendrogram – 3 wk ahead inc death – FL



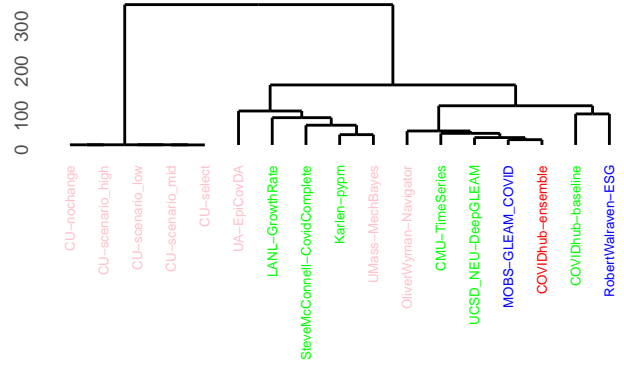
Dendrogram – 4 wk ahead inc death – FL



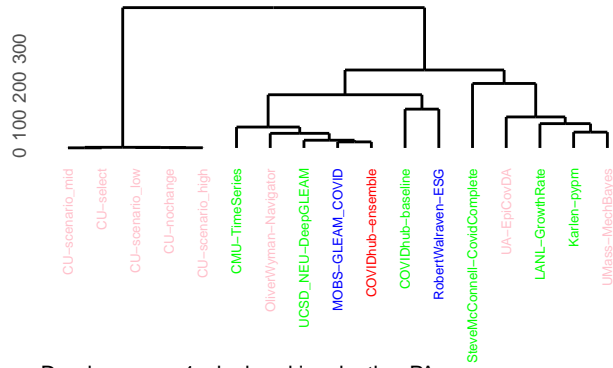
Dendrogram – 1 wk ahead inc death – NY



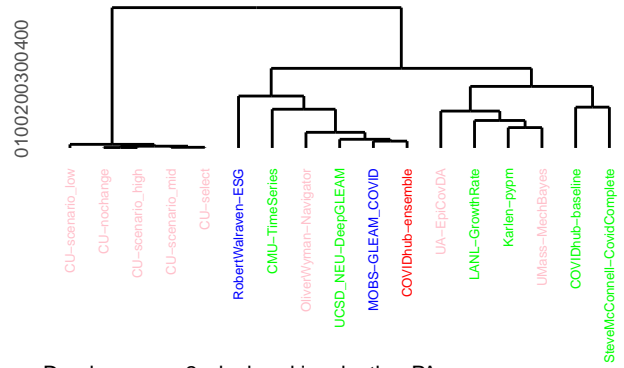
Dendrogram – 2 wk ahead inc death – NY



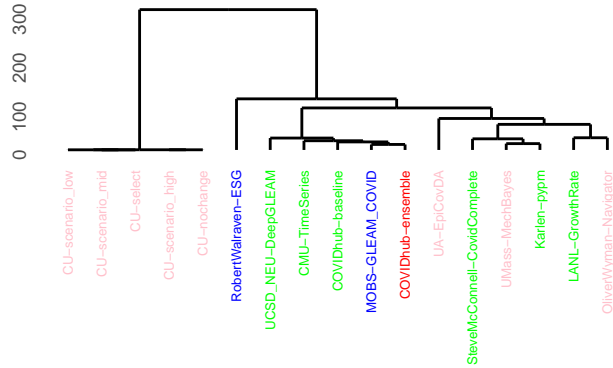
Dendrogram – 3 wk ahead inc death – NY



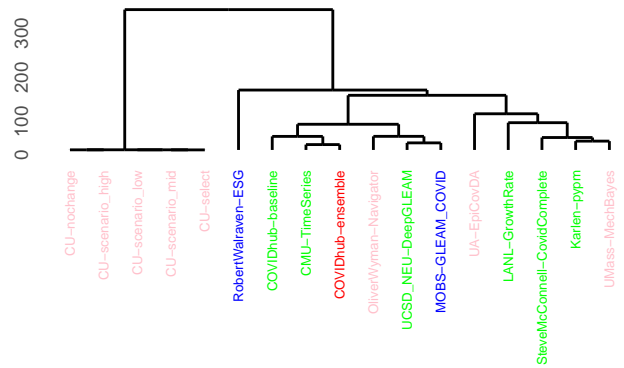
Dendrogram – 4 wk ahead inc death – NY



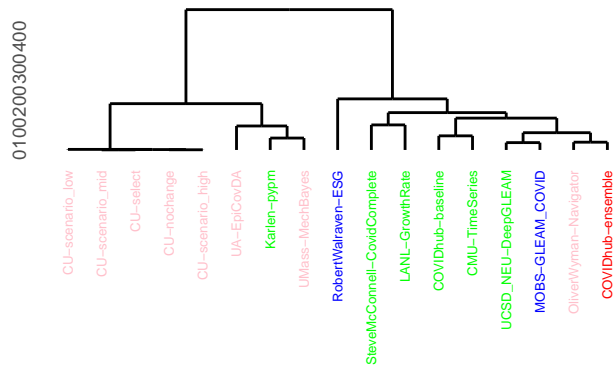
Dendrogram – 1 wk ahead inc death – PA



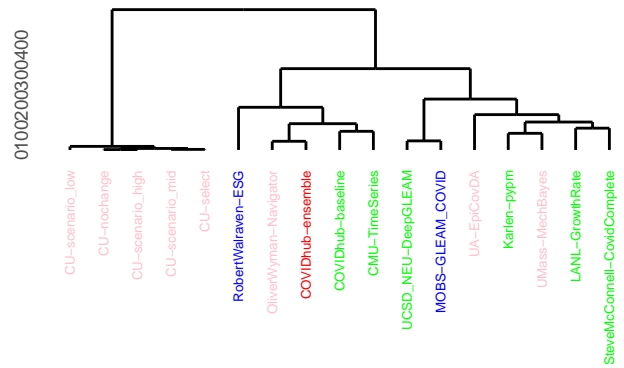
Dendrogram – 2 wk ahead inc death – PA

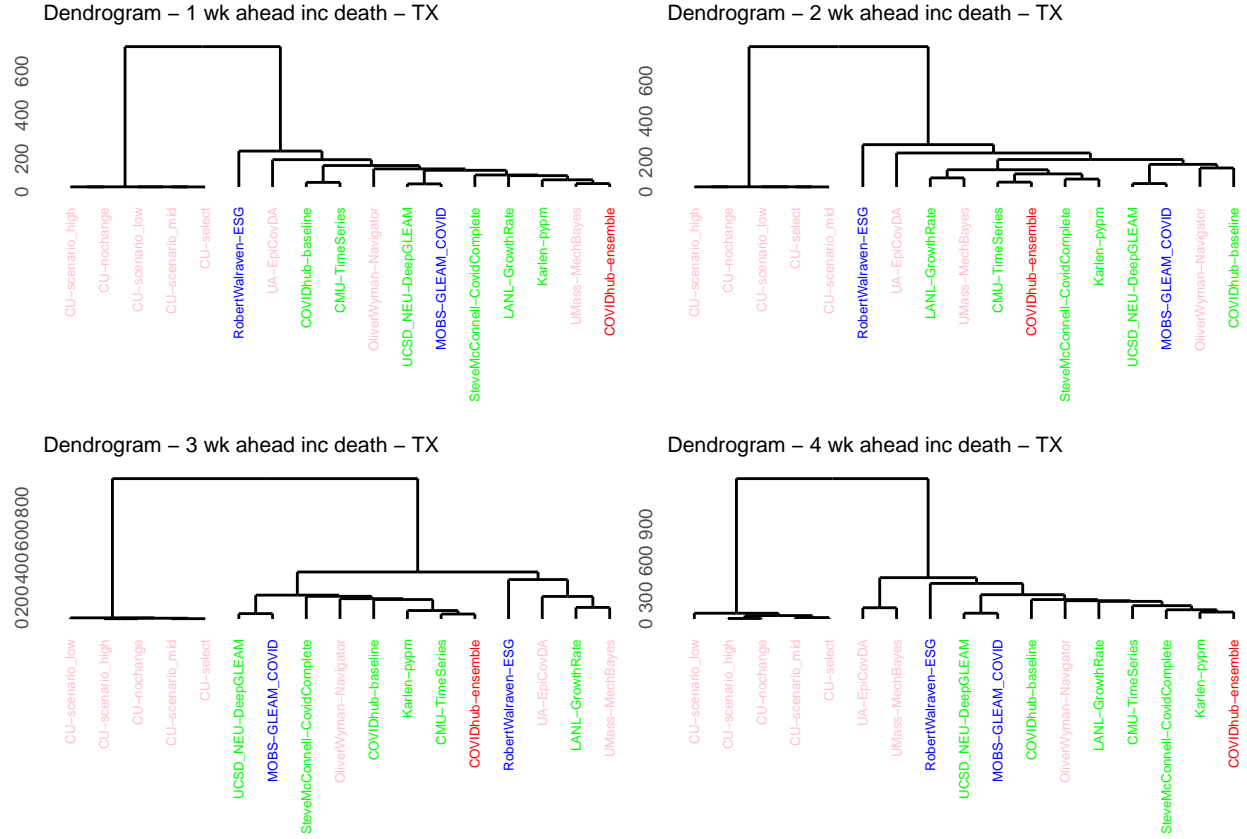


Dendrogram – 3 wk ahead inc death – PA



Dendrogram – 4 wk ahead inc death – PA



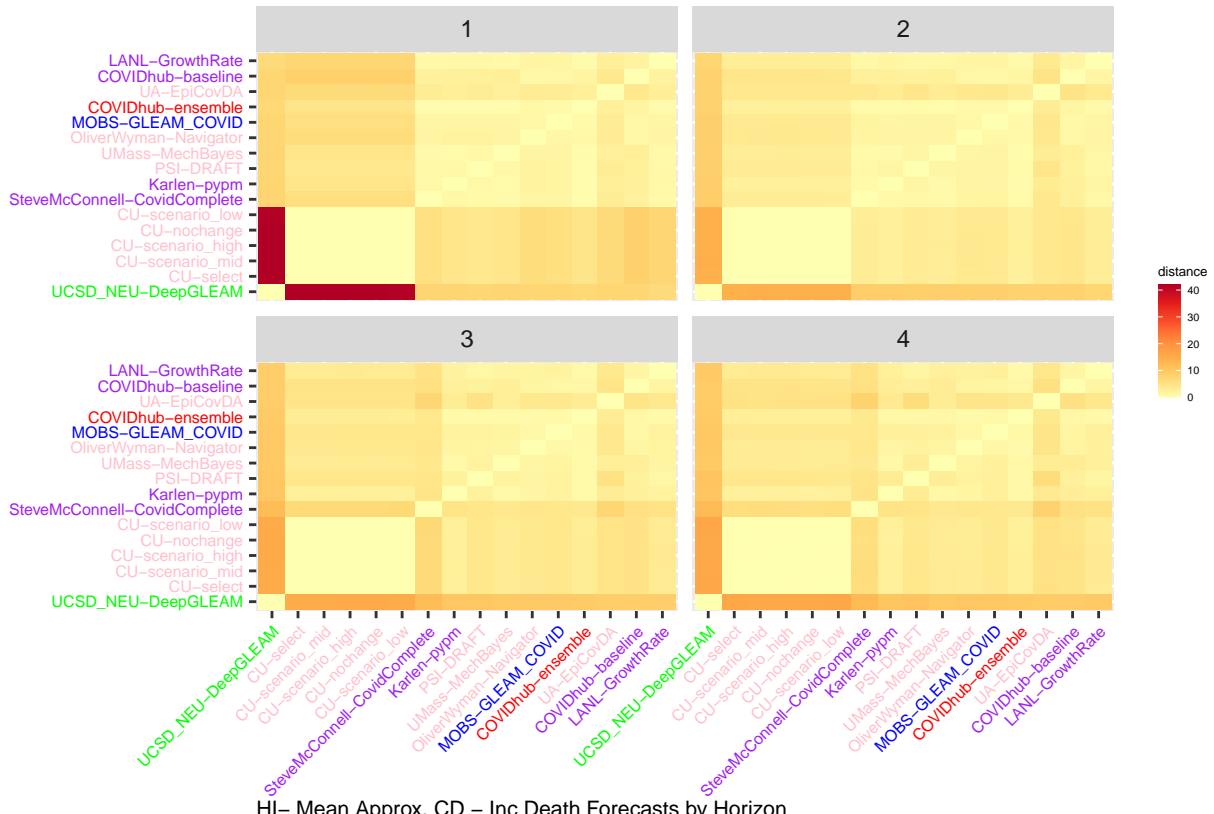


5 locations with the lowest number of COVID-19 deaths by the end of February 2021

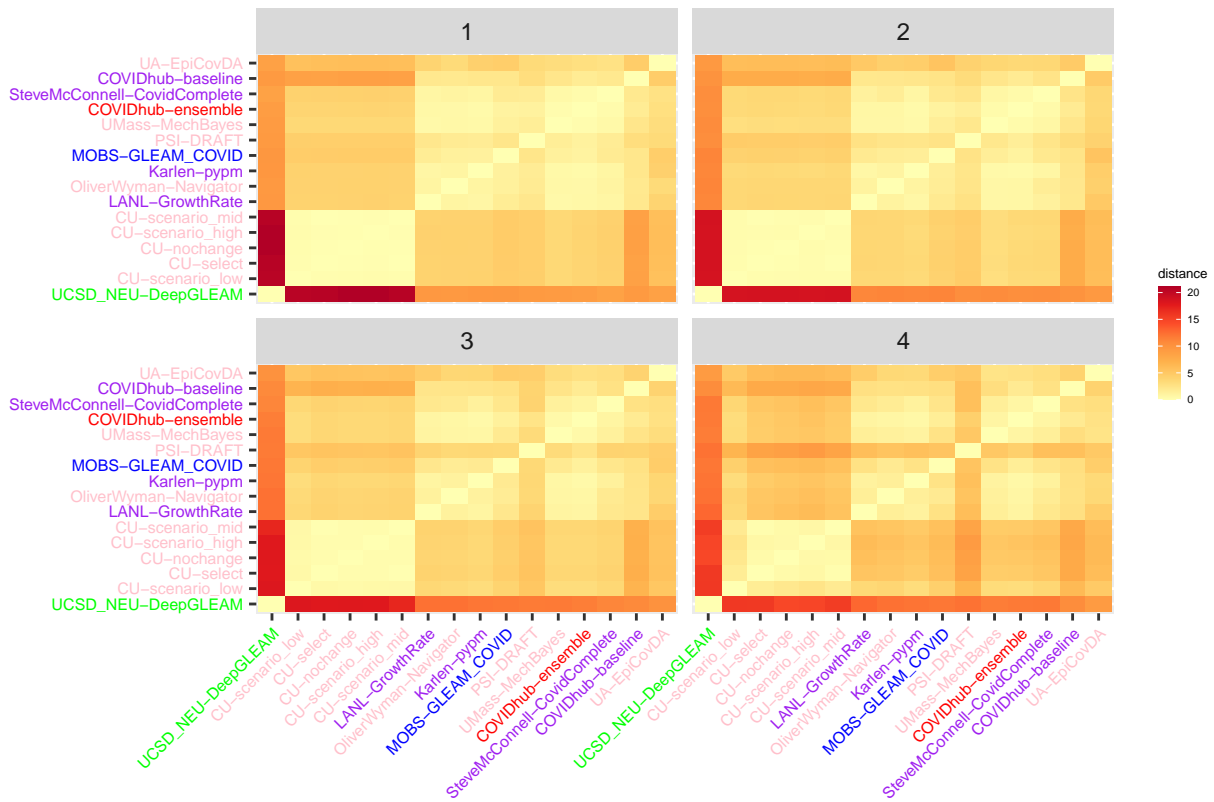
The pairwise approximated Cramer's distances are calculated for the models that have complete submissions for all target, all 5 locations with the lowest number of COVID-19 deaths by the end of February 2021, all probability levels, from the target end date of October, 17th 2020 to May 29th, 2021 .

We can visualize the mean approximated pairwise distances across all time points in a heat map shown below. 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.

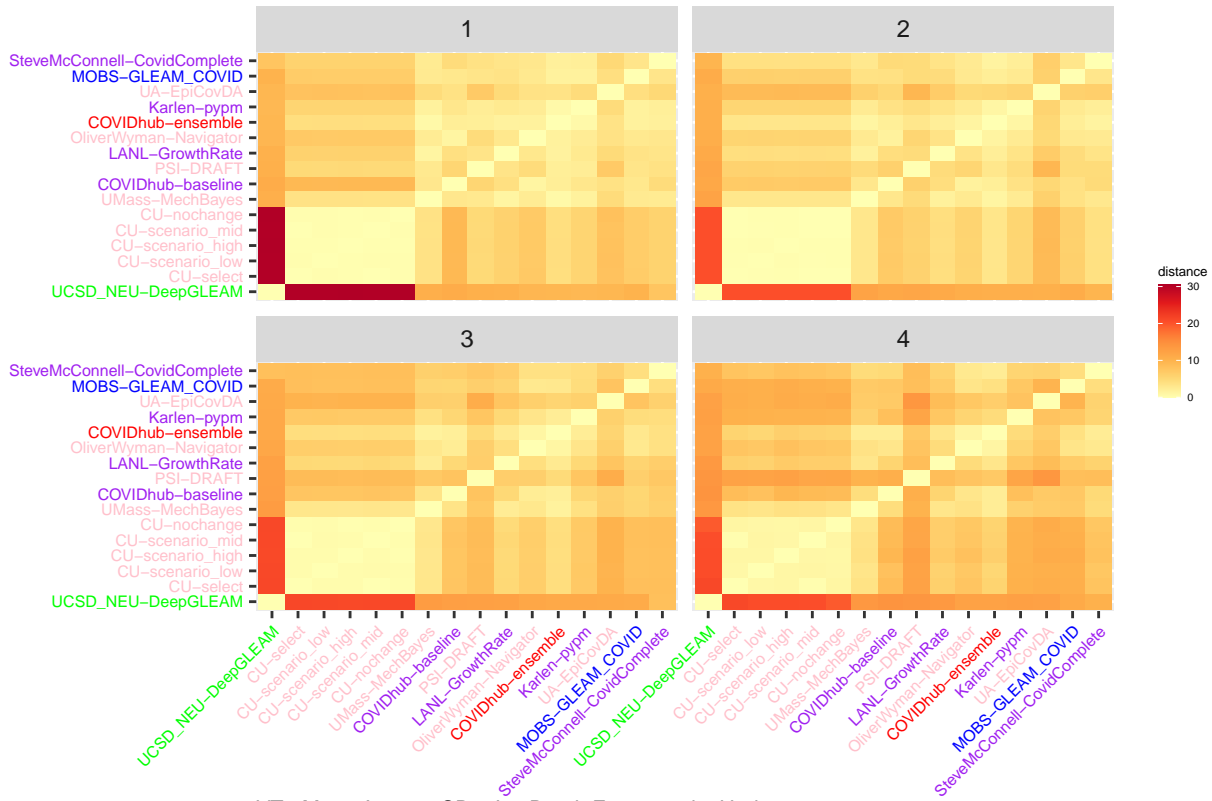
AK– Mean Approx. CD – Inc Death Forecasts by Horizon



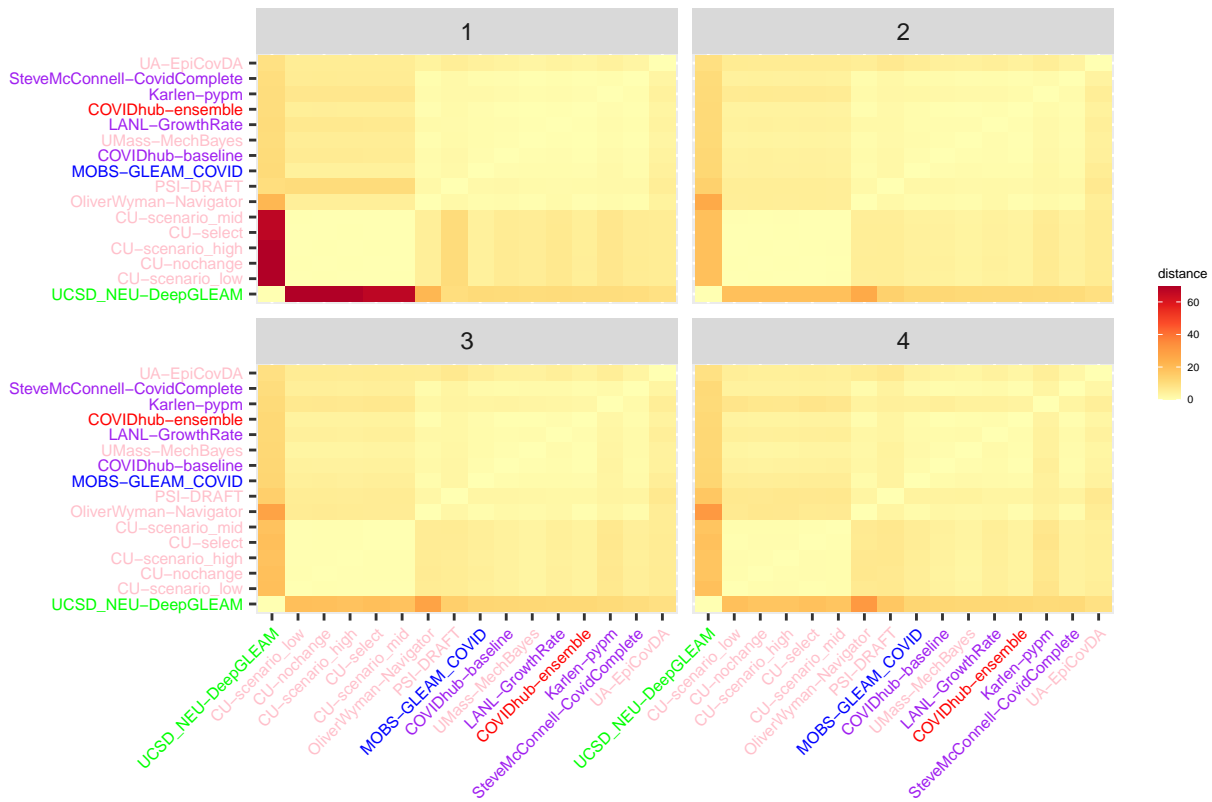
HI– Mean Approx. CD – Inc Death Forecasts by Horizon

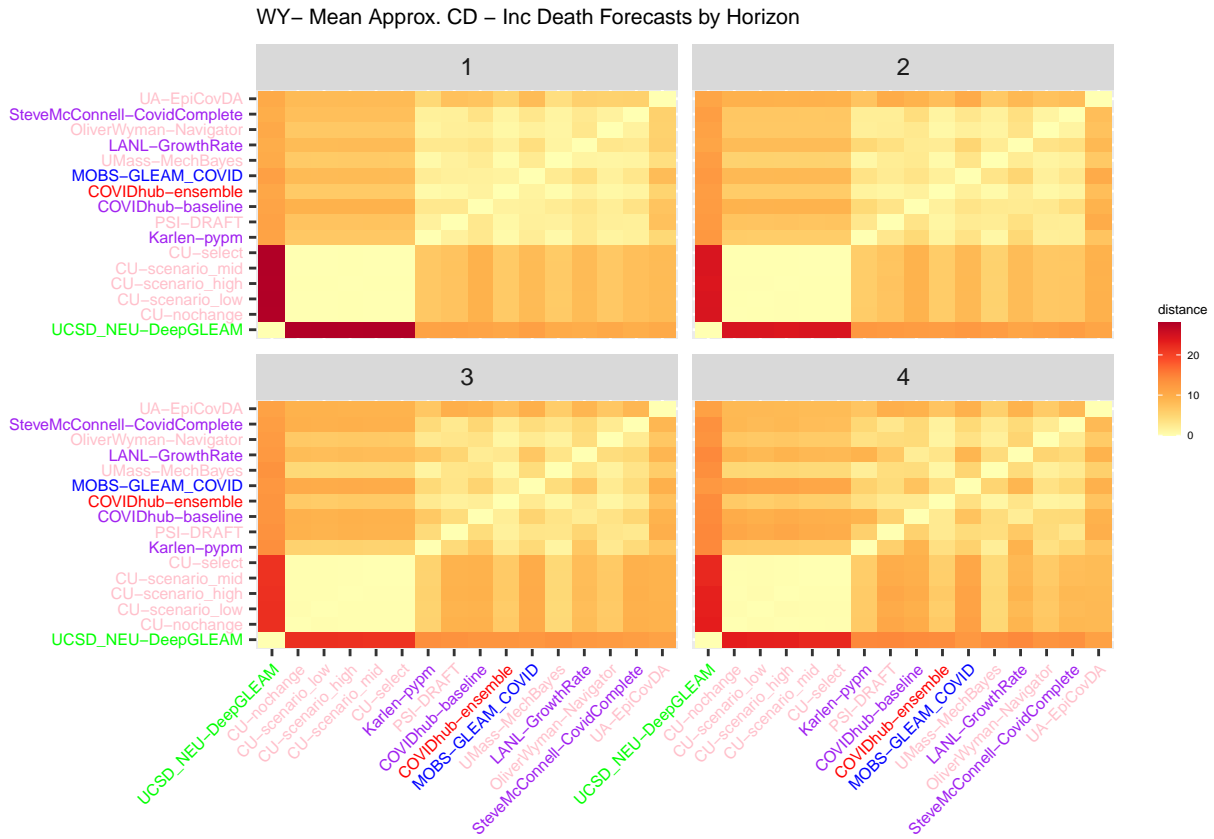


ME– Mean Approx. CD – Inc Death Forecasts by Horizon



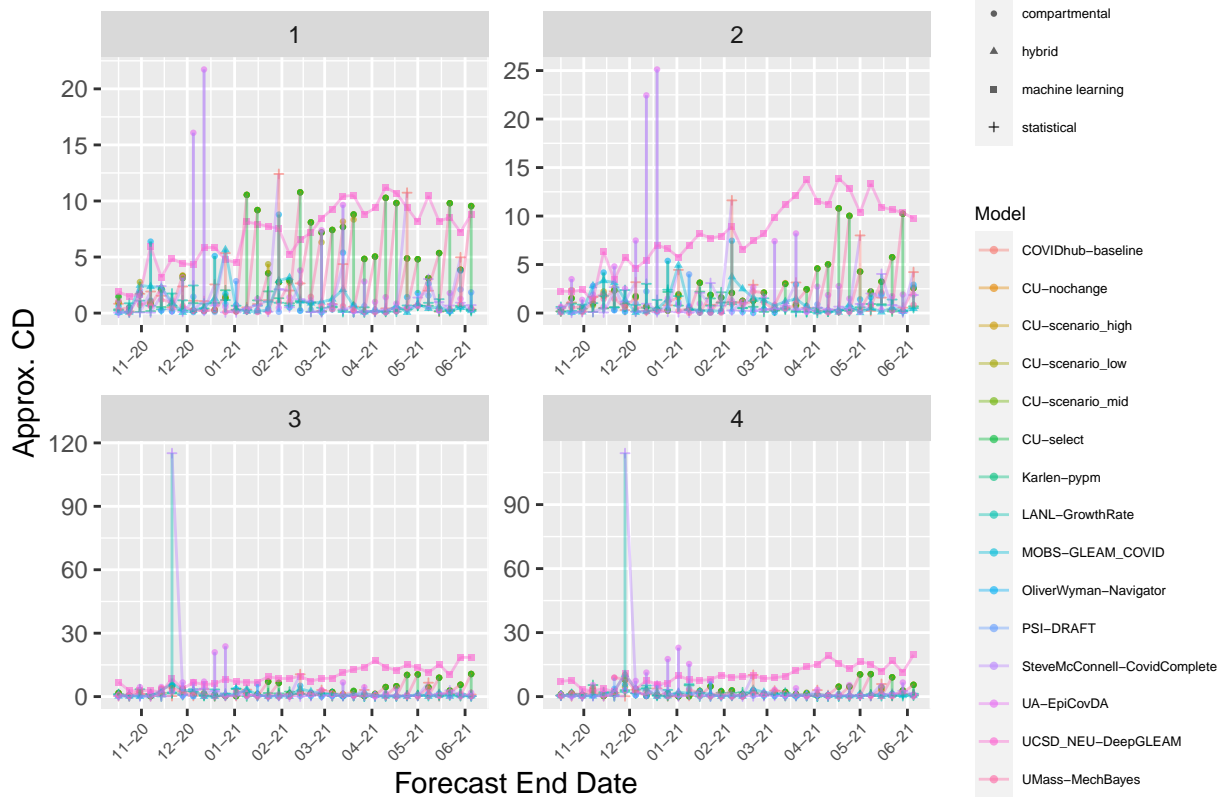
VT– Mean Approx. CD – Inc Death Forecasts by Horizon



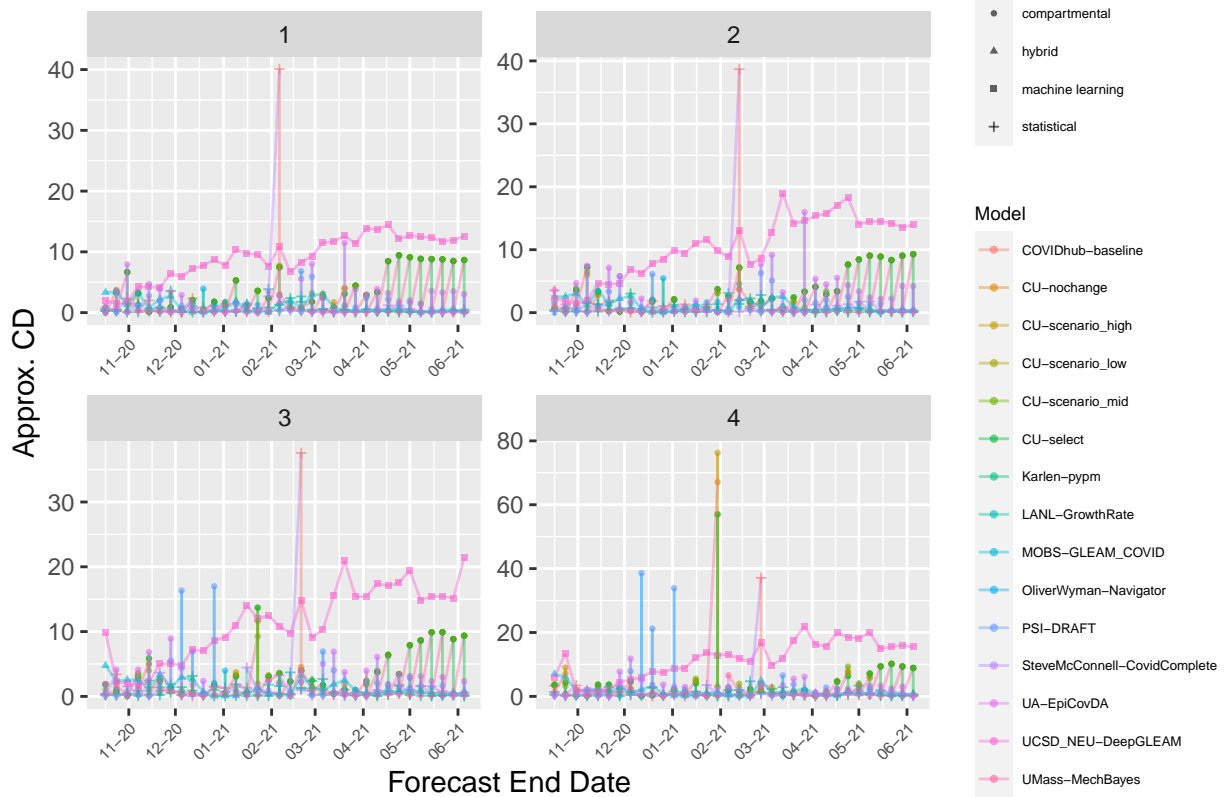


We can also look at the approximated pairwise distances to see how the models become more similar or dissimilar over time.

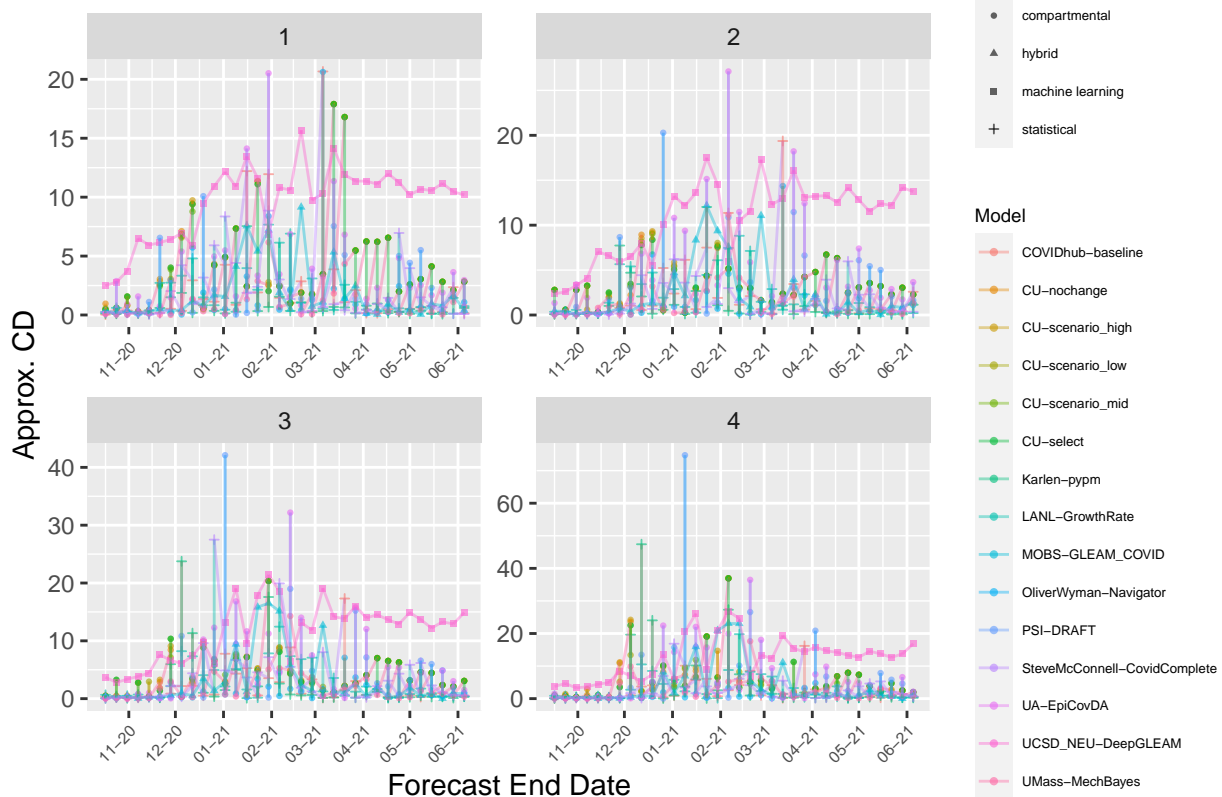
Approx. CD from COVIDhub-ensemble Over Time – AK



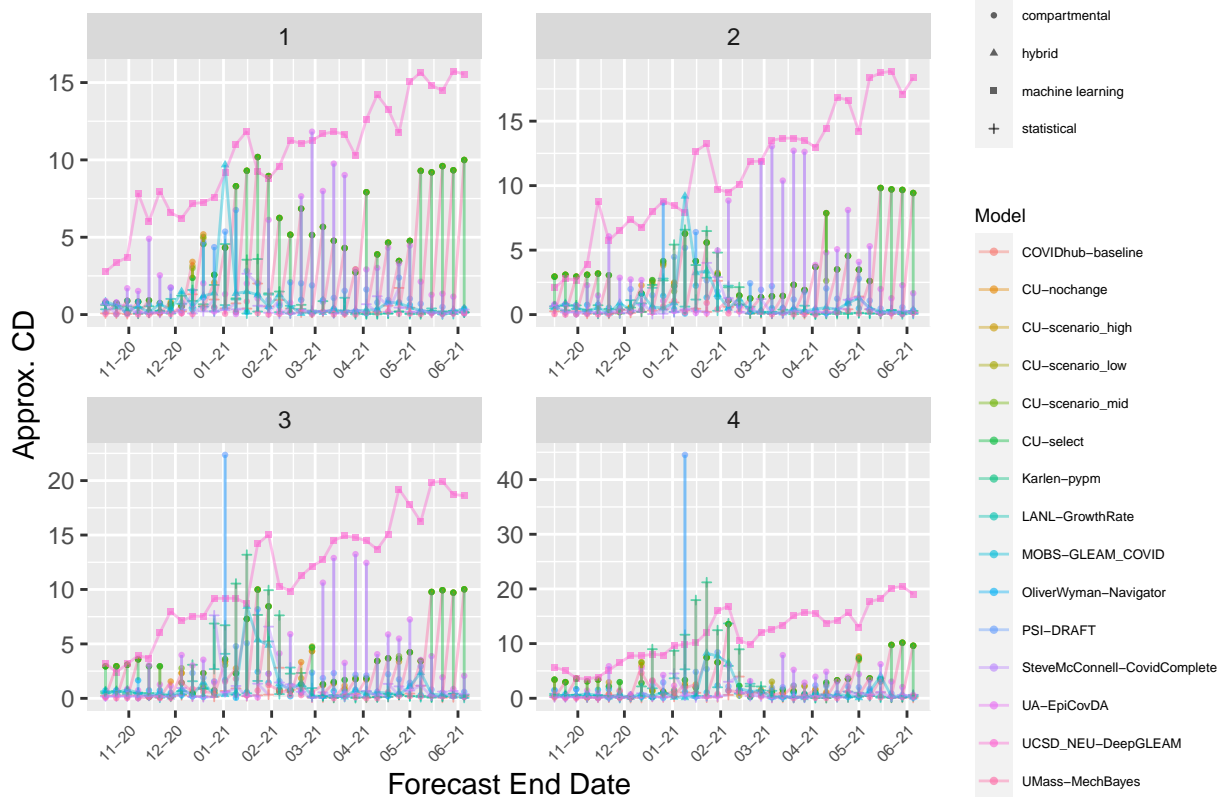
Approx. CD from COVIDhub-ensemble Over Time – HI

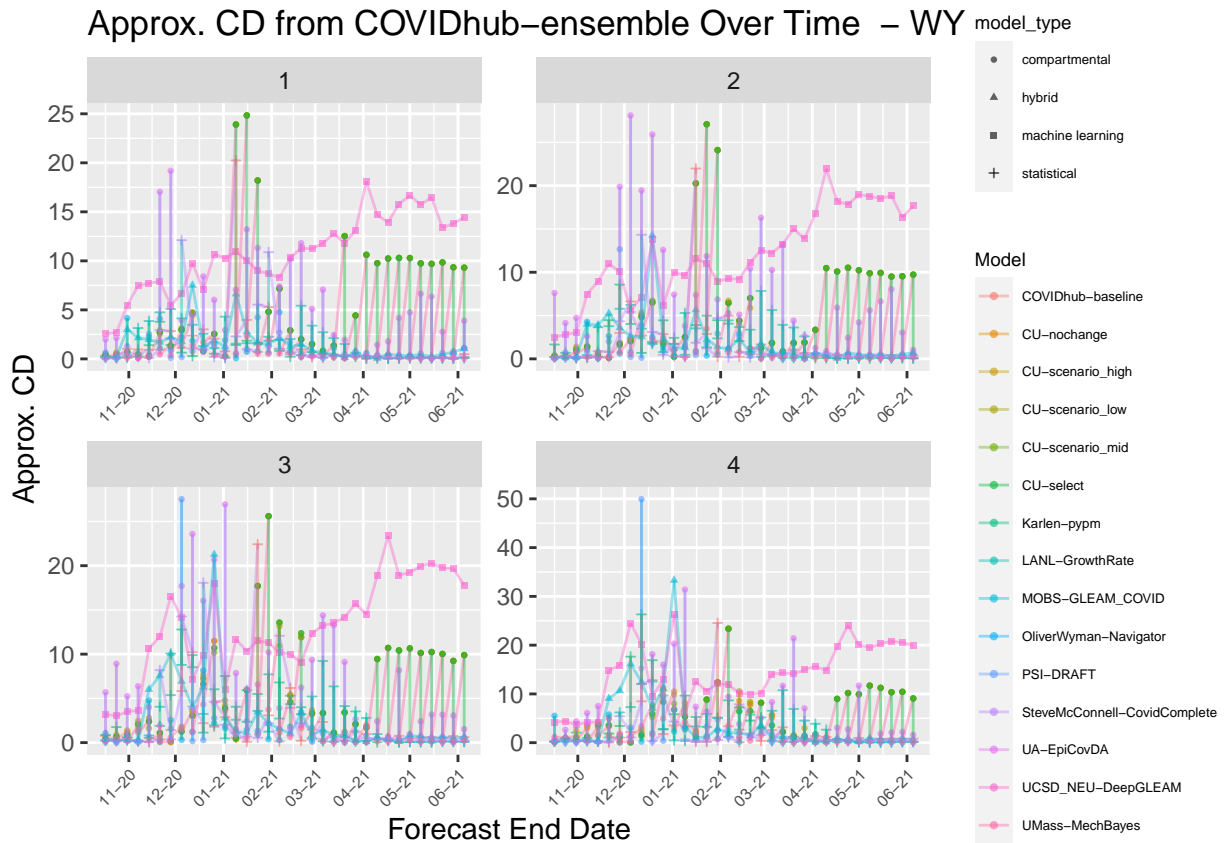


Approx. CD from COVIDhub-ensemble Over Time – ME



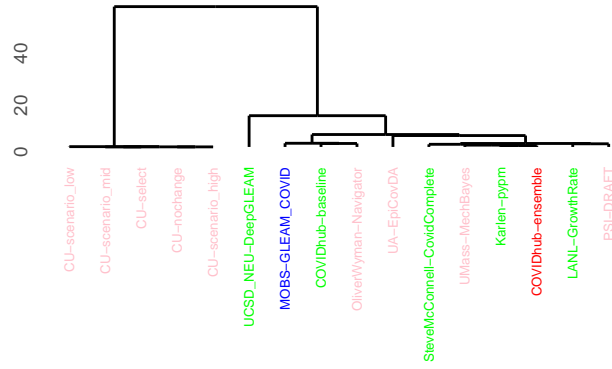
Approx. CD from COVIDhub-ensemble Over Time – VT



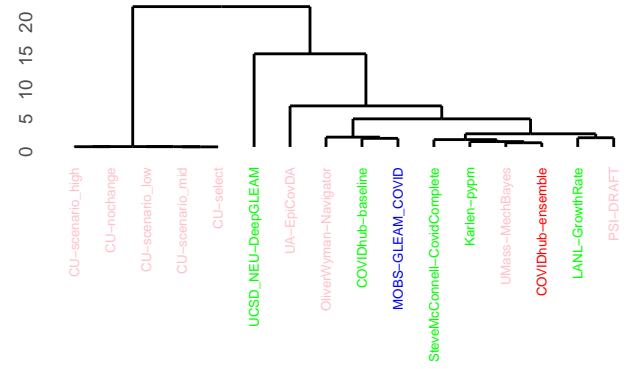


We can cluster the distances using hierarchical clustering. Different linkages will result in different clusters, we probably should investigate more later.

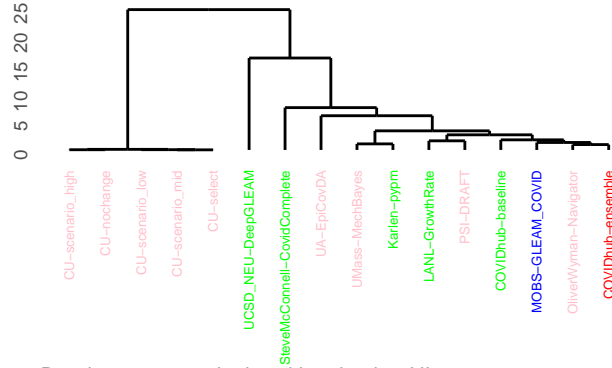
Dendrogram – 1 wk ahead inc death – AK



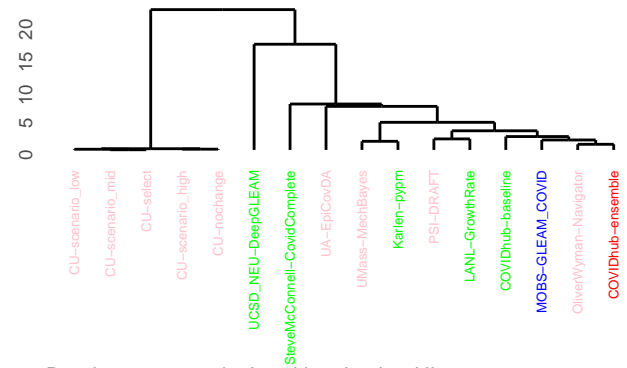
Dendrogram – 2 wk ahead inc death – AK



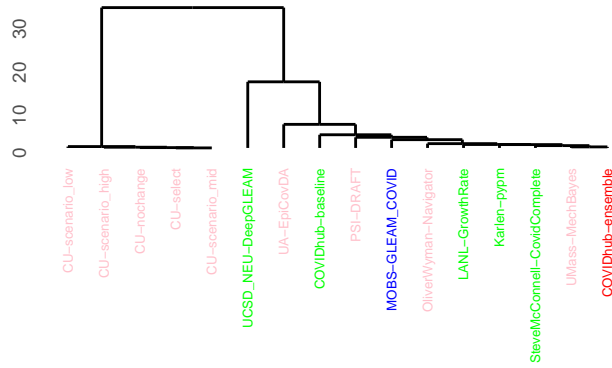
Dendrogram – 3 wk ahead inc death – AK



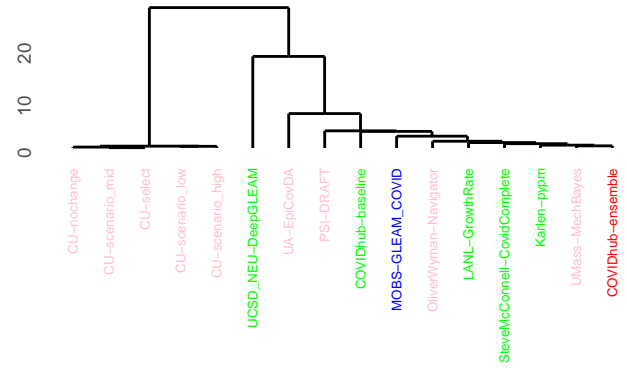
Dendrogram – 4 wk ahead inc death – AK



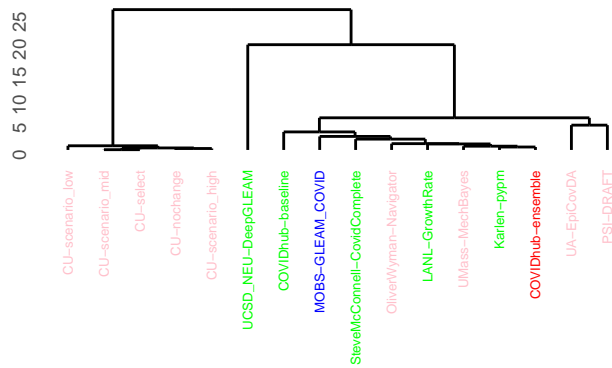
Dendrogram – 1 wk ahead inc death – HI



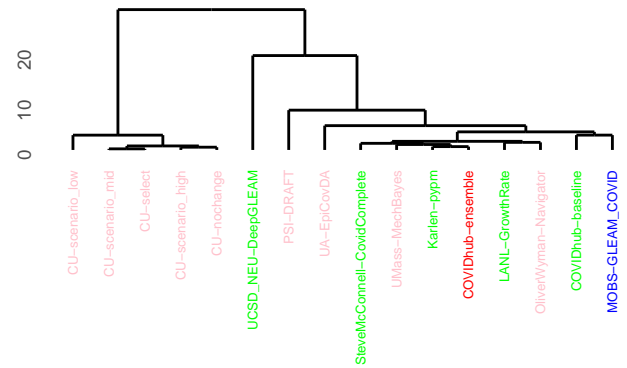
Dendrogram – 2 wk ahead inc death – HI



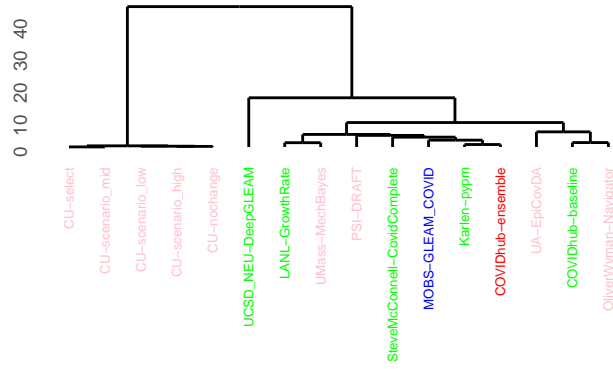
Dendrogram – 3 wk ahead inc death – HI



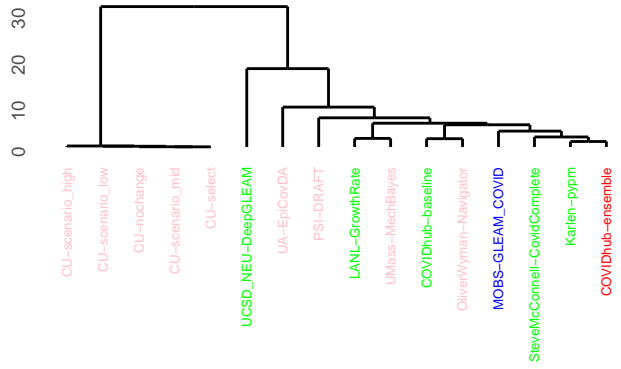
Dendrogram – 4 wk ahead inc death – HI



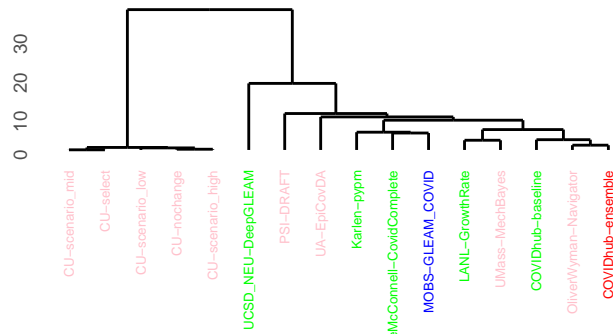
Dendrogram – 1 wk ahead inc death – ME



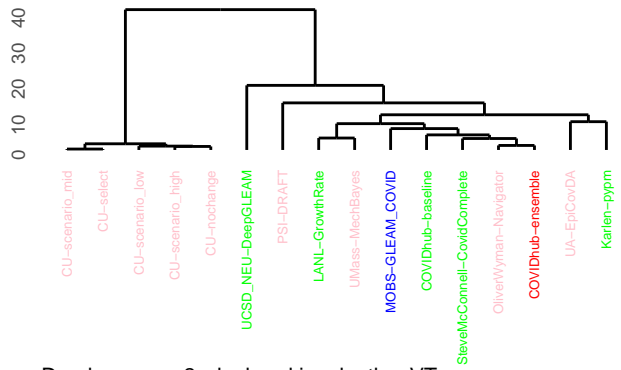
Dendrogram – 2 wk ahead inc death – ME



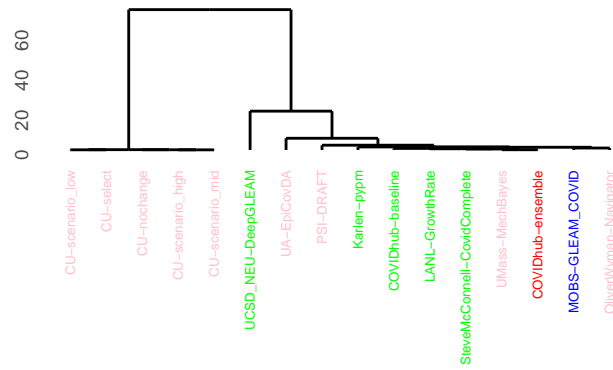
Dendrogram – 3 wk ahead inc death – ME



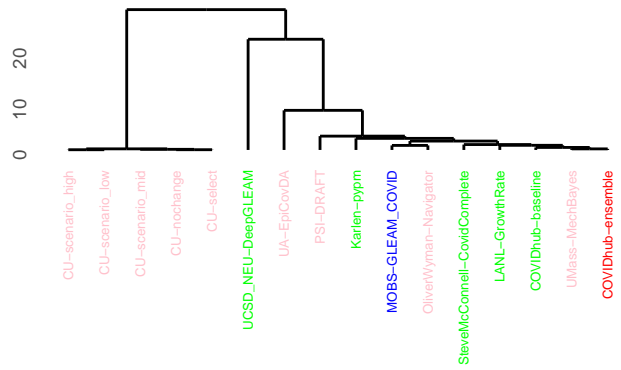
Dendrogram – 4 wk ahead inc death – ME



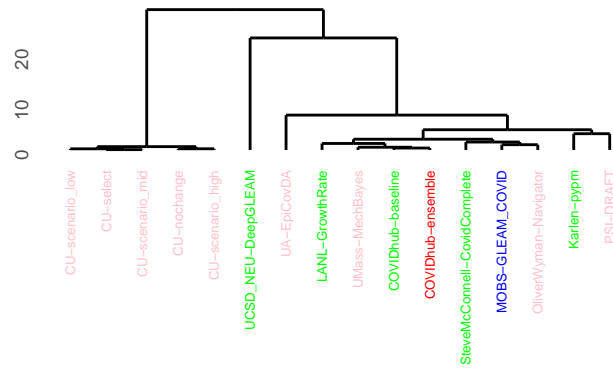
Dendrogram – 1 wk ahead inc death – VT



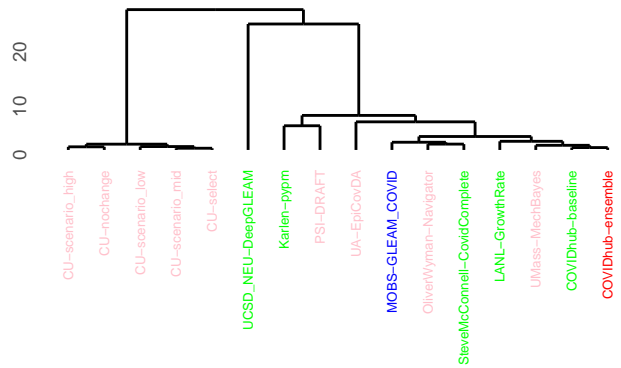
Dendrogram – 2 wk ahead inc death – VT

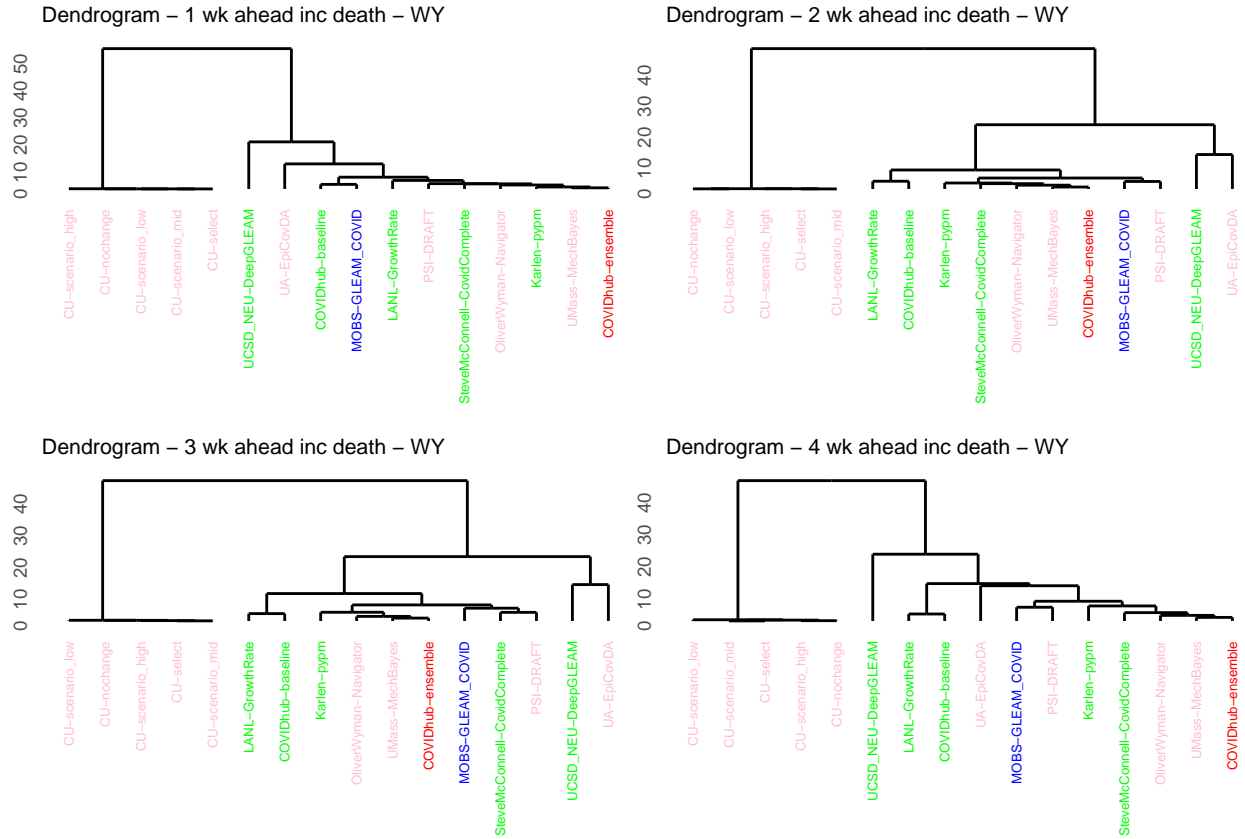


Dendrogram – 3 wk ahead inc death – VT



Dendrogram – 4 wk ahead inc death – VT





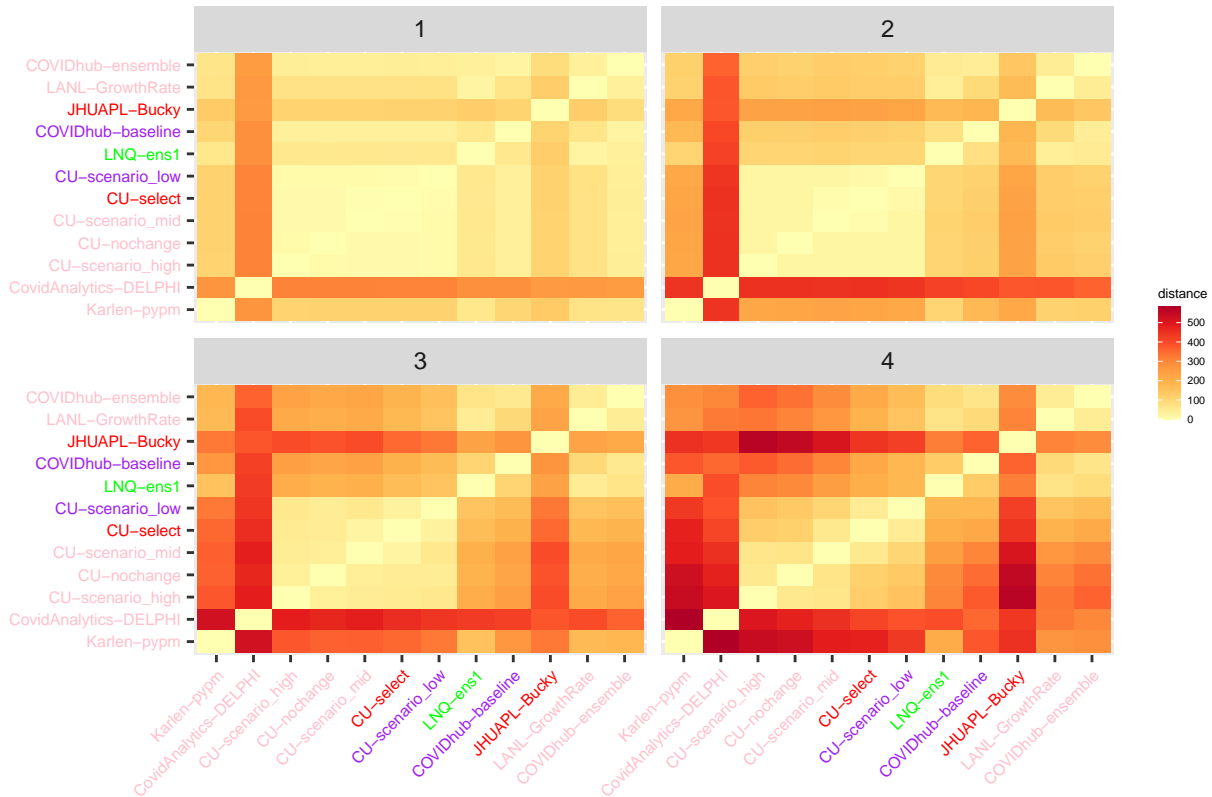
COVID-19 Forecasting Model Similarity Analysis for 1-4 Week Ahead Incident Case

5 locations with the highest number of COVID-19 cumulative cases from 01/25/2020 to 02/07/2021

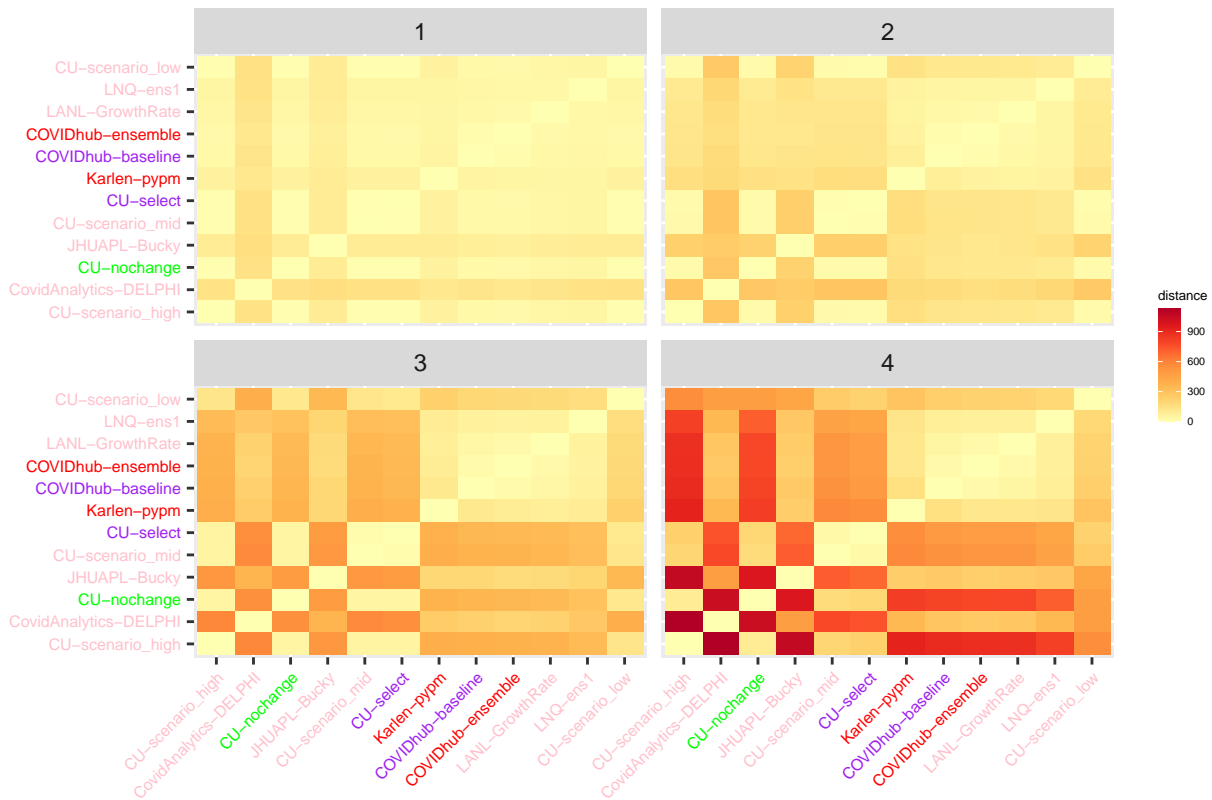
The pairwise approximated Cramer's distances are calculated for the models that have complete submissions for all target, all probability levels, from mid-October 2020 until May 24th, 2021 for 5 locations with the highest cumulative cases.

Below are the heat maps of mean approximated pairwise CD across time by location-target:

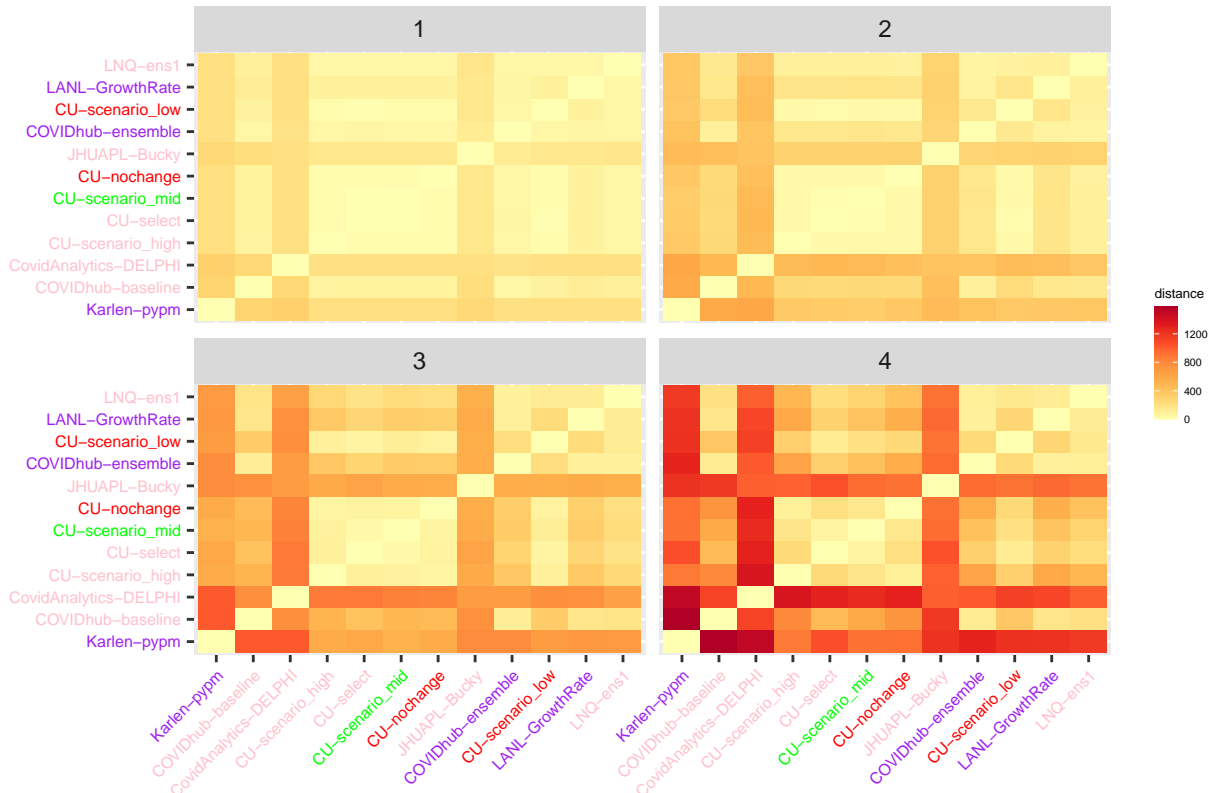
DC– Mean Approx. CD – Inc Case Forecasts by Horizon



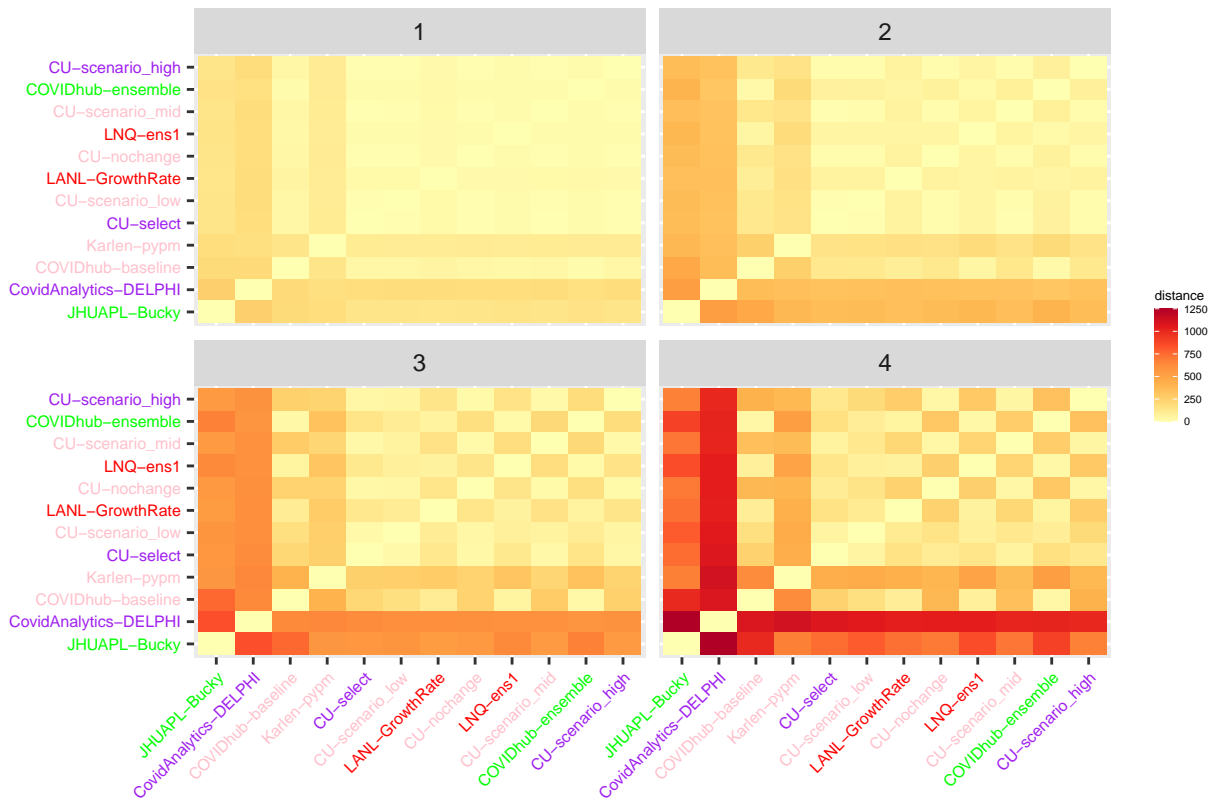
HI– Mean Approx. CD – Inc Case Forecasts by Horizon



ME– Mean Approx. CD – Inc Case Forecasts by Horizon



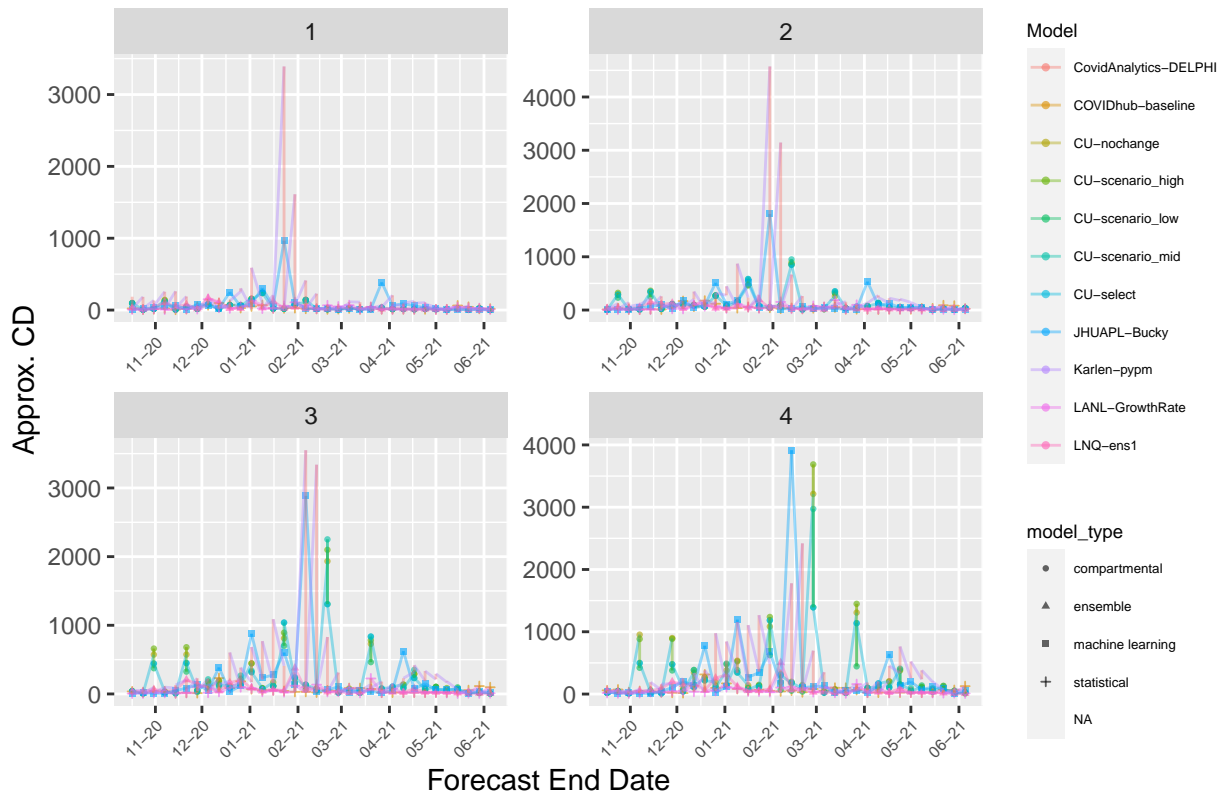
VT– Mean Approx. CD – Inc Case Forecasts by Horizon



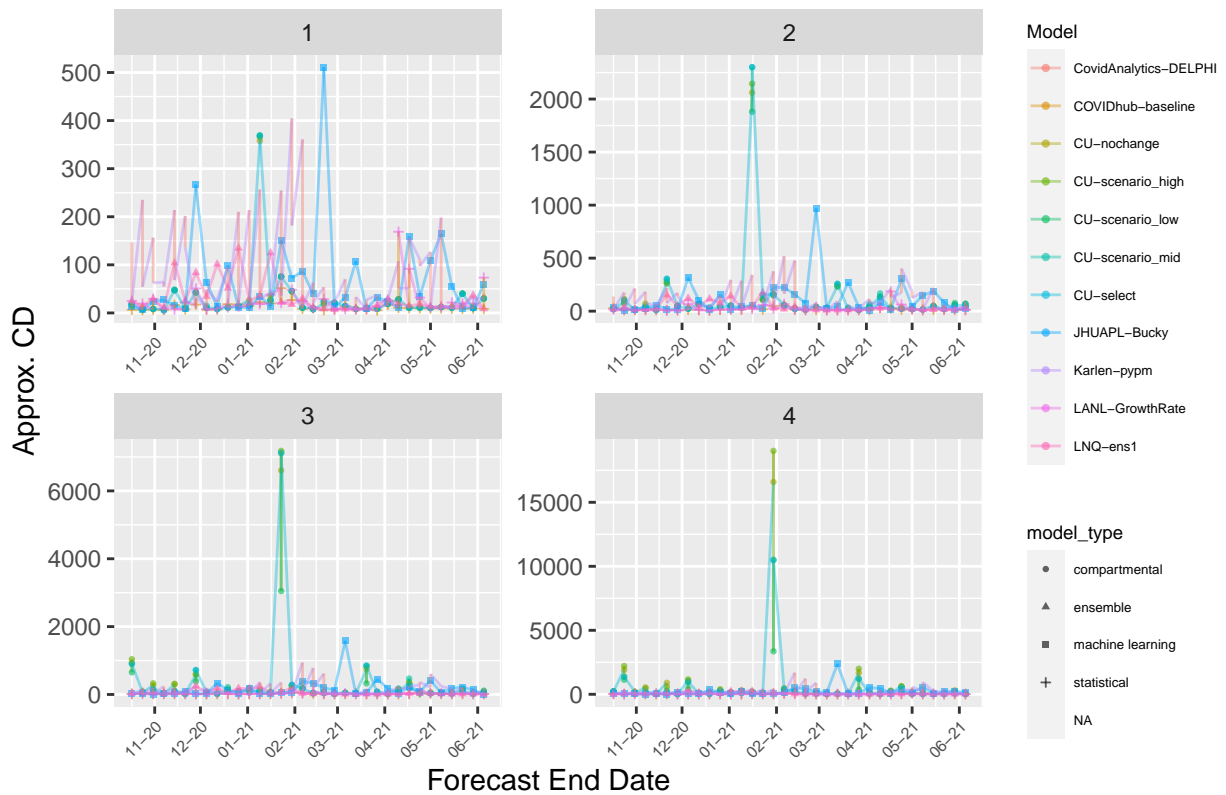


Below are the plots of approximated pairwise distances over time. Unlike indicent death targets, there are too many model pairs to show (12 models for inc death and 17 models for inc case, so we have a combination of 6 choose 2 with no repeats (15) vs a combination of 11 choose 2 with no repeats (55)). So, I only pick the first 15 of the combinations:

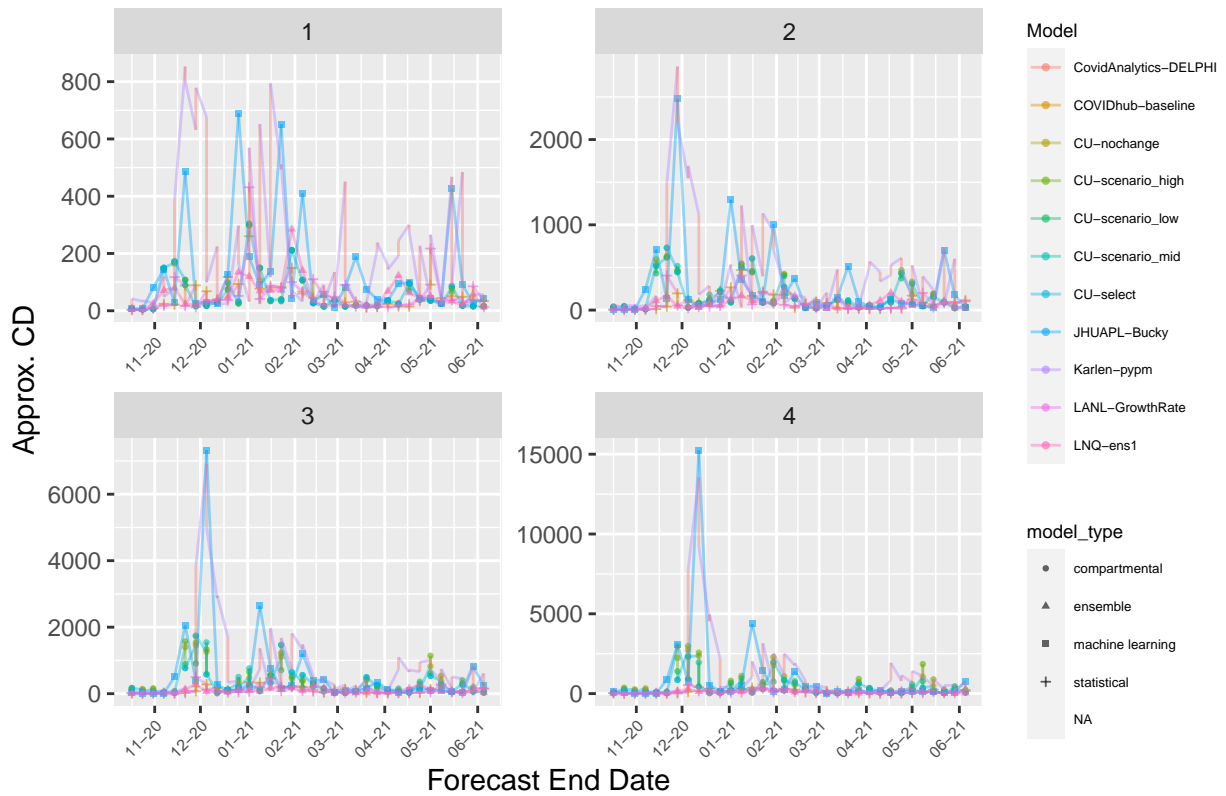
Approx. CD from COVIDhub-ensemble Over Time – DC



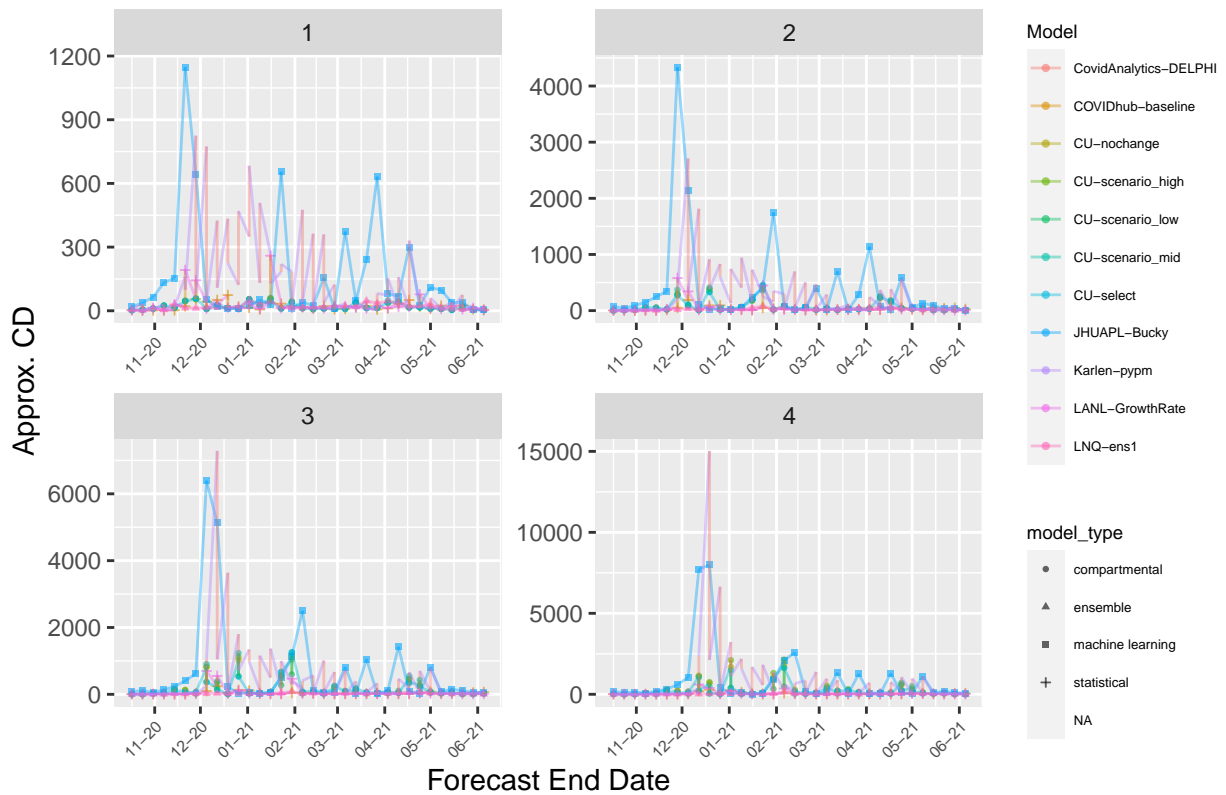
Approx. CD from COVIDhub-ensemble Over Time – HI



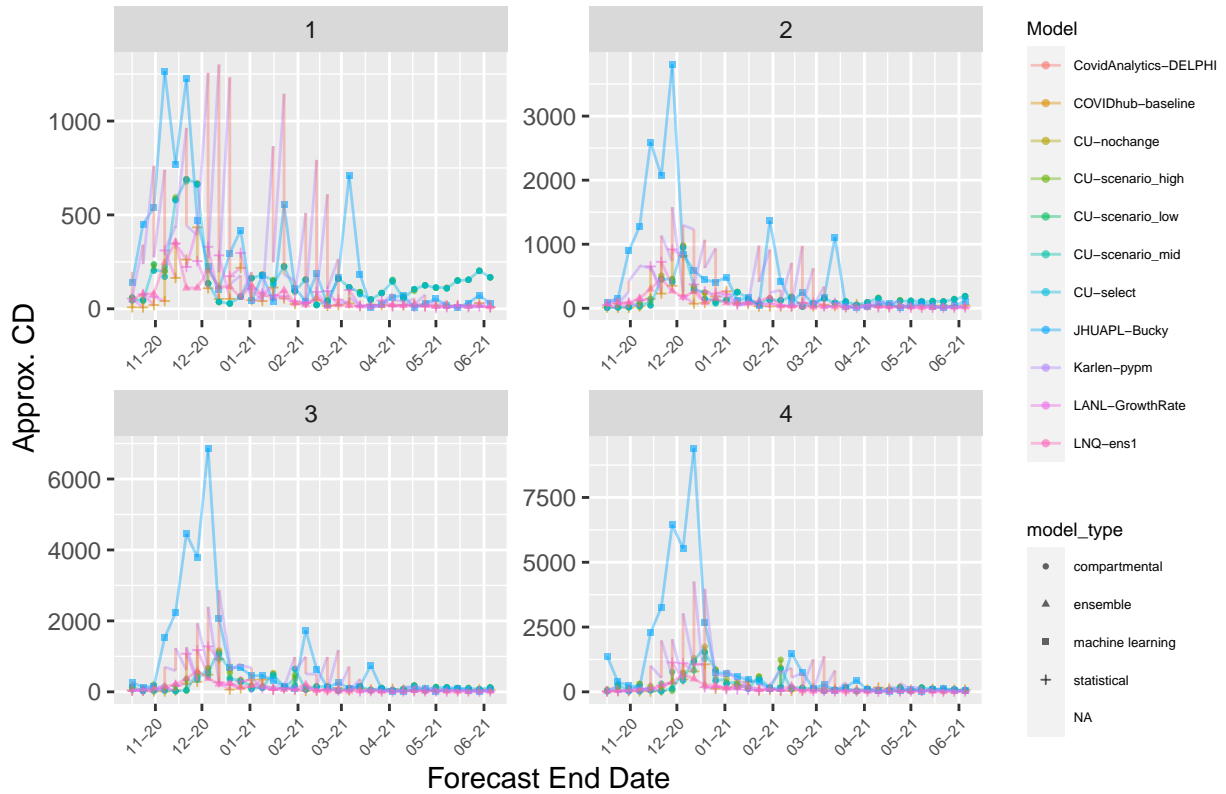
Approx. CD from COVIDhub-ensemble Over Time – ME



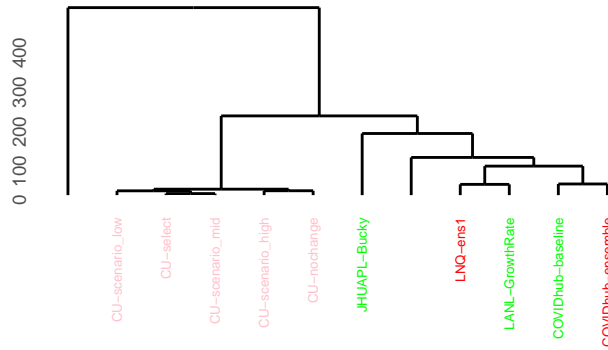
Approx. CD from COVIDhub-ensemble Over Time – VT



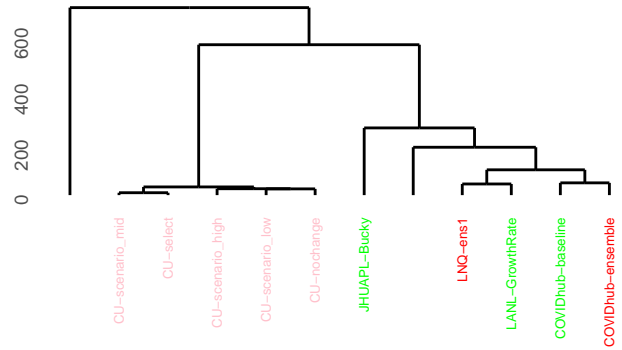
Approx. CD from COVIDhub-ensemble Over Time – WY



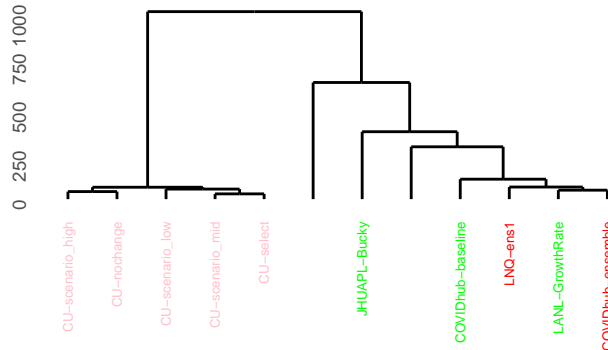
Dendrogram – 1 wk ahead inc case – DC



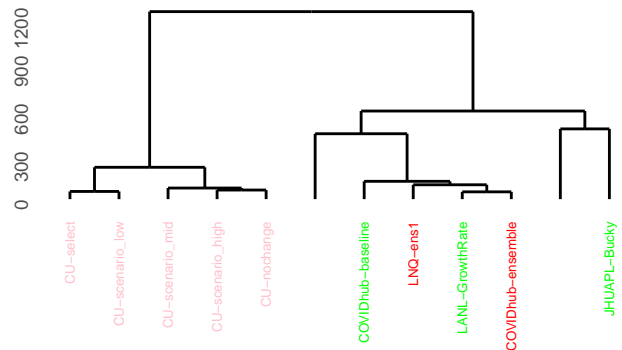
Dendrogram – 2 wk ahead inc case – DC



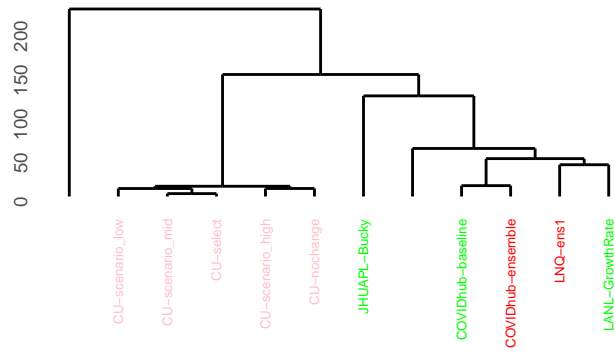
Dendrogram – 3 wk ahead inc case – DC



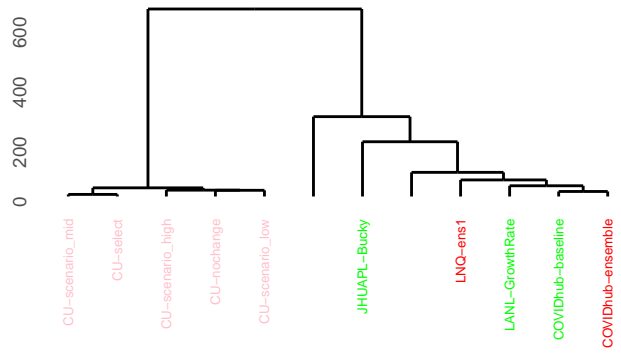
Dendrogram – 4 wk ahead inc case – DC



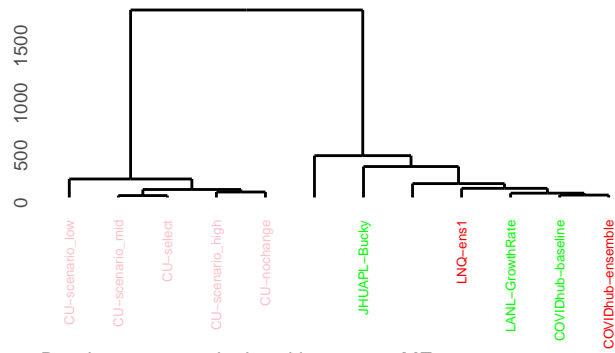
Dendrogram – 1 wk ahead inc case – HI



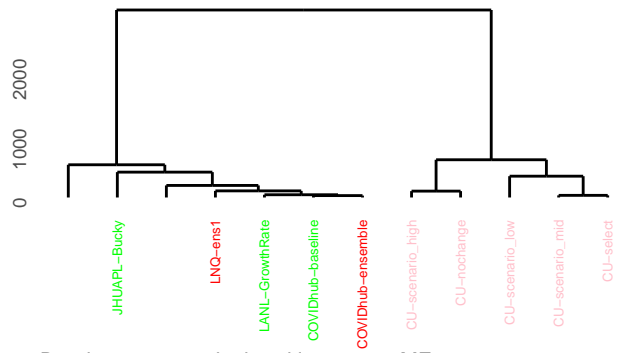
Dendrogram – 2 wk ahead inc case – HI



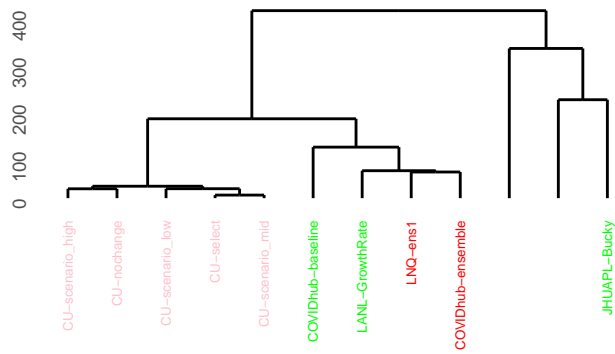
Dendrogram – 3 wk ahead inc case – HI



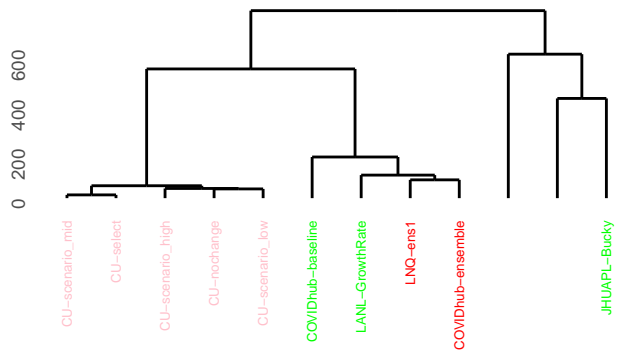
Dendrogram – 4 wk ahead inc case – HI



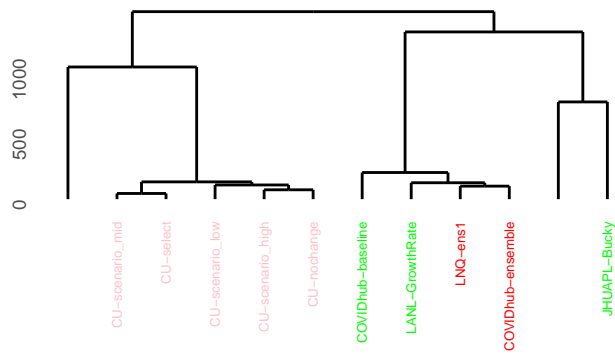
Dendrogram – 1 wk ahead inc case – ME



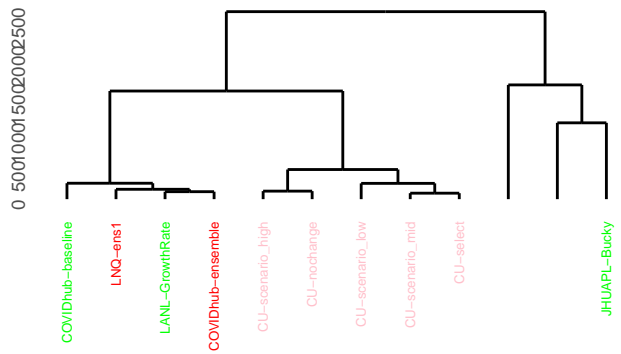
Dendrogram – 2 wk ahead inc case – ME



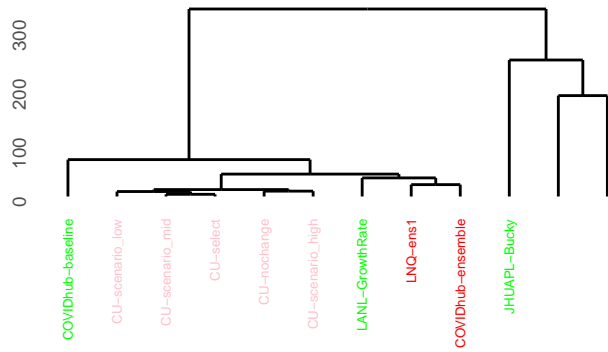
Dendrogram – 3 wk ahead inc case – ME



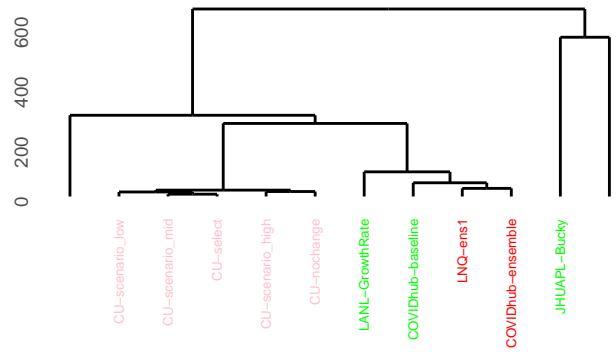
Dendrogram – 4 wk ahead inc case – ME



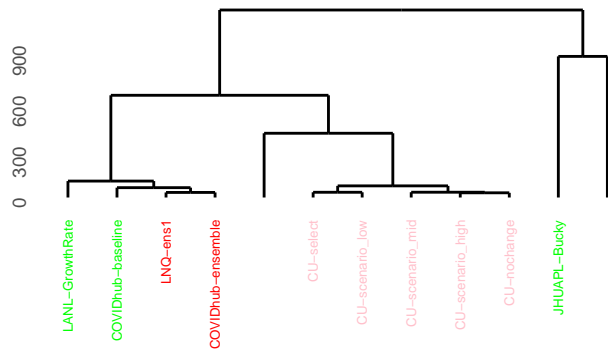
Dendrogram – 1 wk ahead inc case – VT



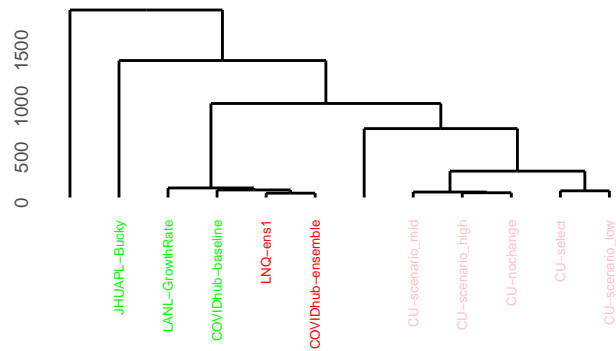
Dendrogram – 2 wk ahead inc case – VT



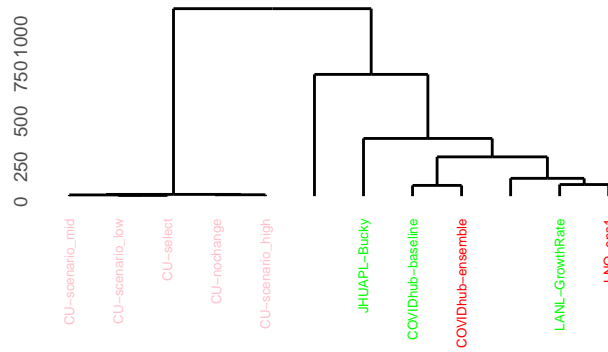
Dendrogram – 3 wk ahead inc case – VT



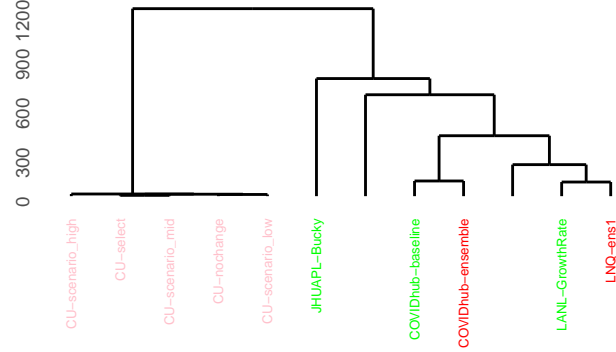
Dendrogram – 4 wk ahead inc case – VT



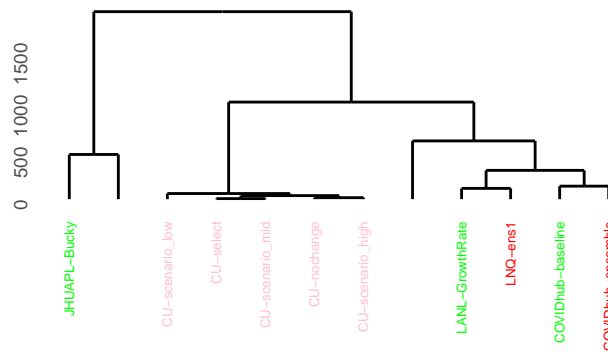
Dendrogram – 1 wk ahead inc case – WY



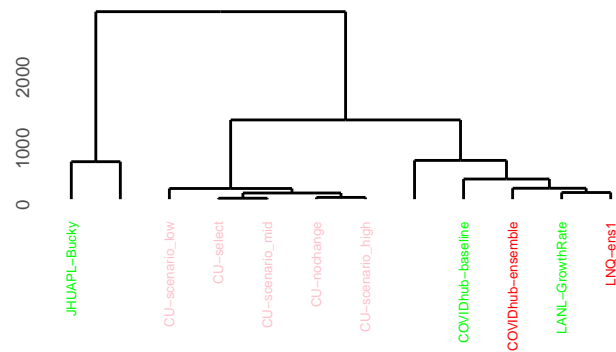
Dendrogram – 2 wk ahead inc case – WY



Dendrogram – 3 wk ahead inc case – WY



Dendrogram – 4 wk ahead inc case – WY



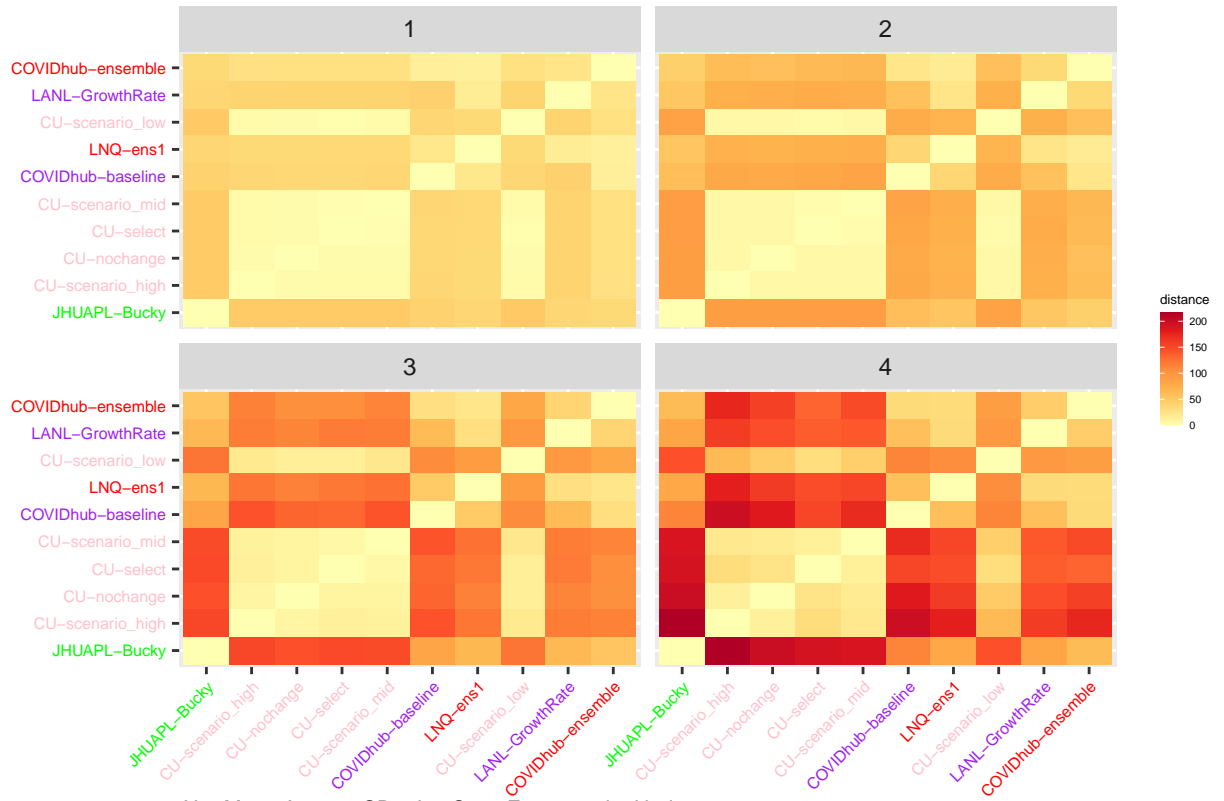
5 locations with the lowest number of COVID-19 cases from 01/25/2020 to 02/07/2021

The pairwise approximated Cramer's distances are calculated for the models that have complete submissions for all target, all probability levels, from mid-October 2020 until May 24th, 2021 for 5 locations with the lowest cumulative cases.

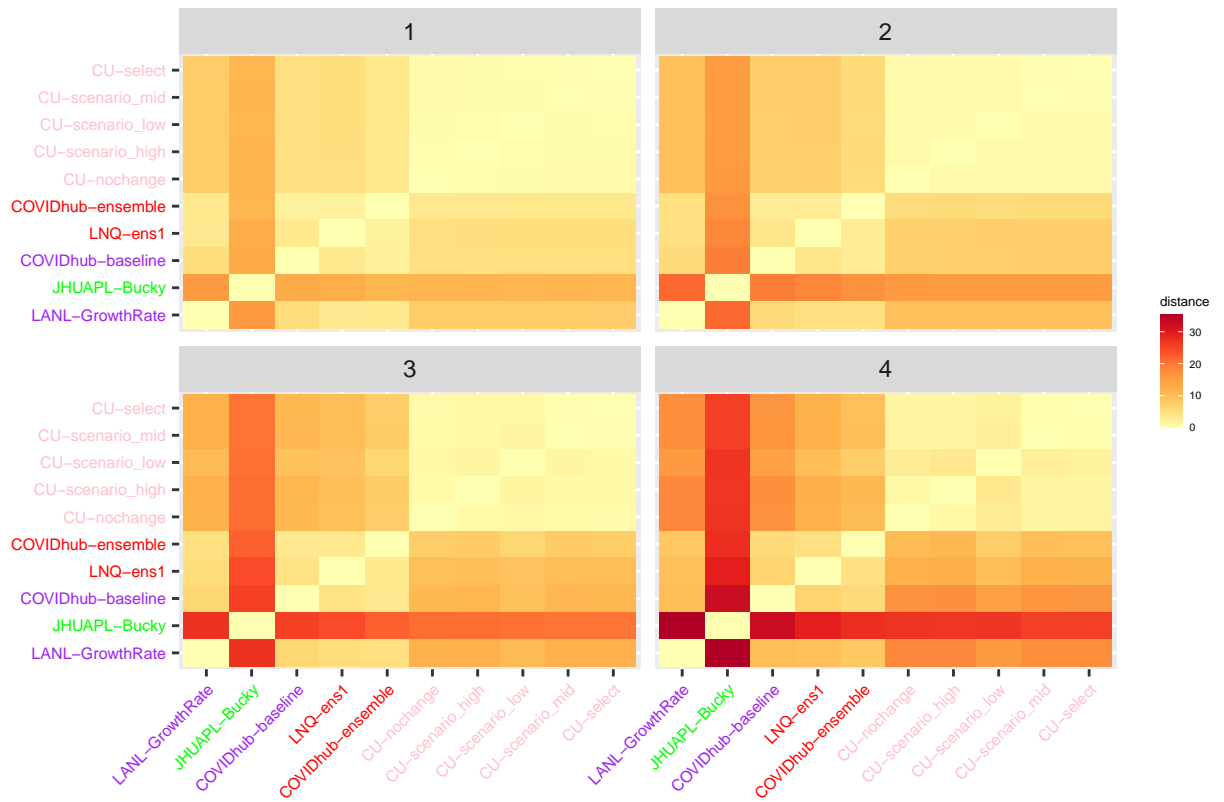
Below are the heatmaps of mean approximated pairwise CD across time by location-target:



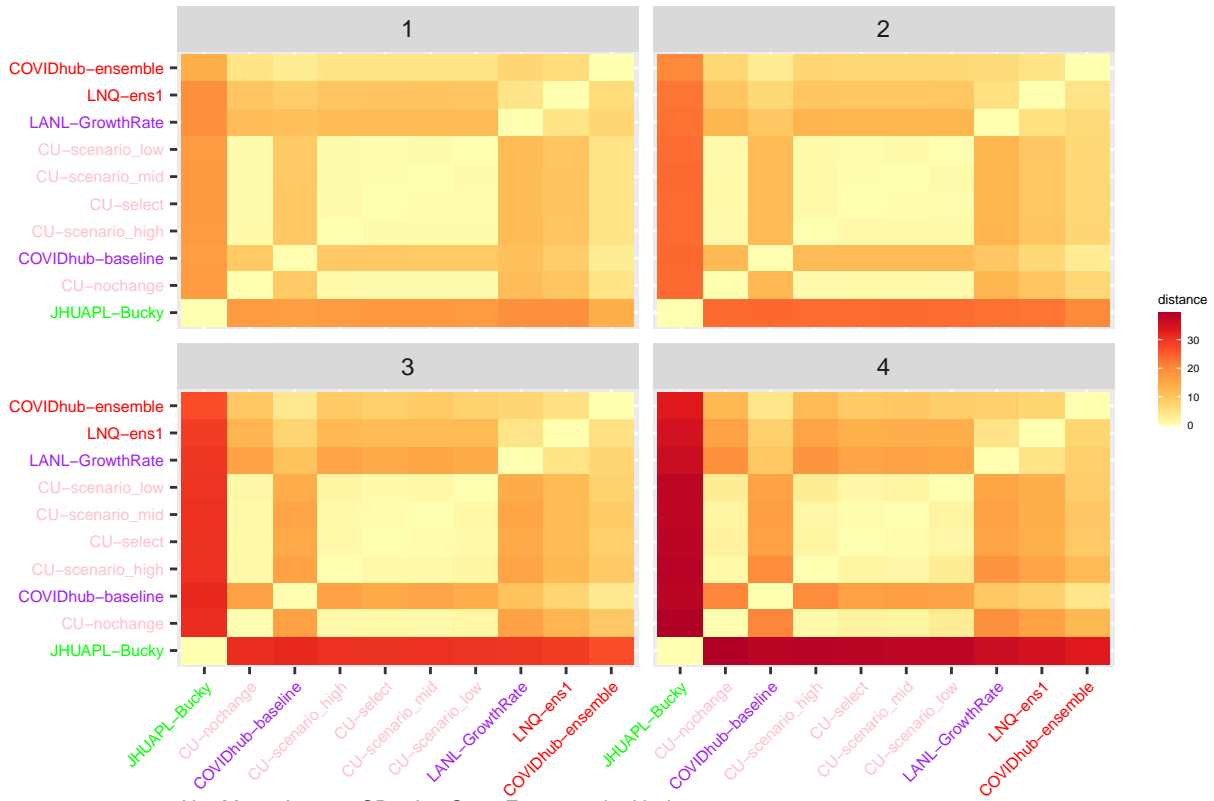
AL– Mean Approx. CD – Inc Case Forecasts by Horizon



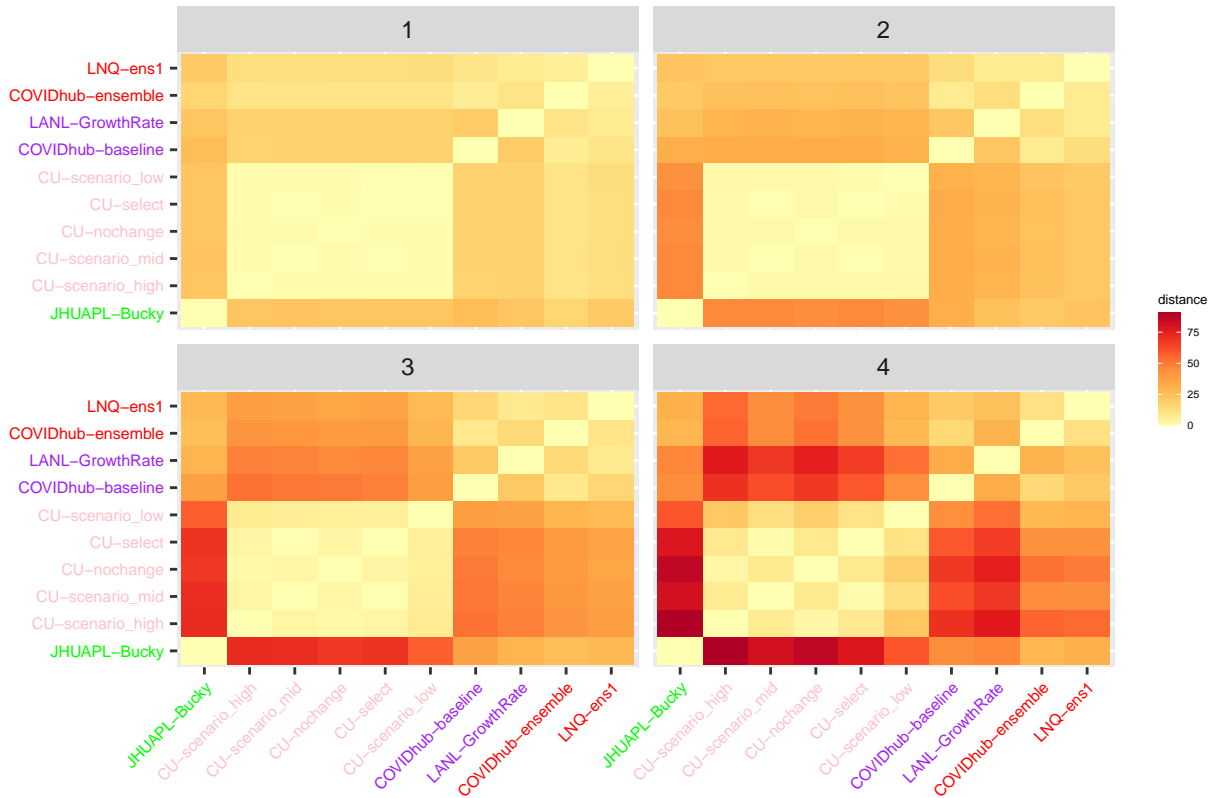
AL– Mean Approx. CD – Inc Case Forecasts by Horizon



AL– Mean Approx. CD – Inc Case Forecasts by Horizon



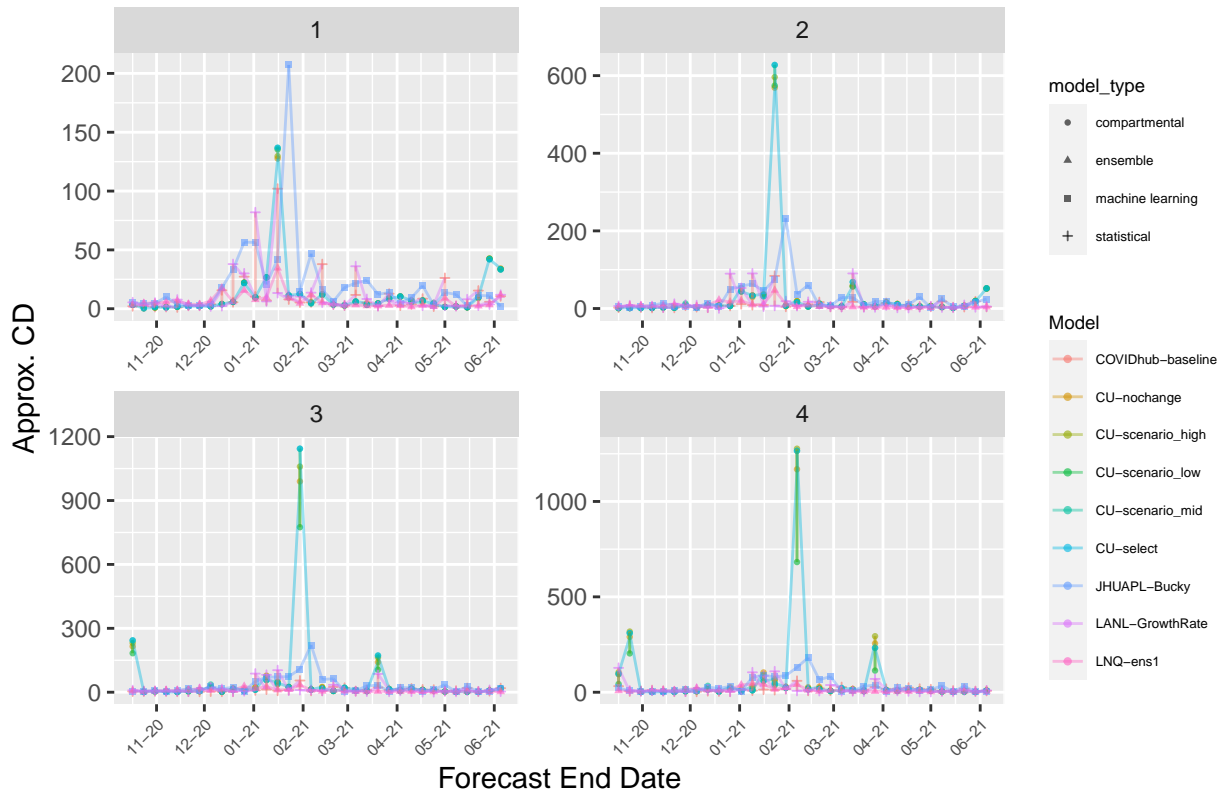
AL– Mean Approx. CD – Inc Case Forecasts by Horizon



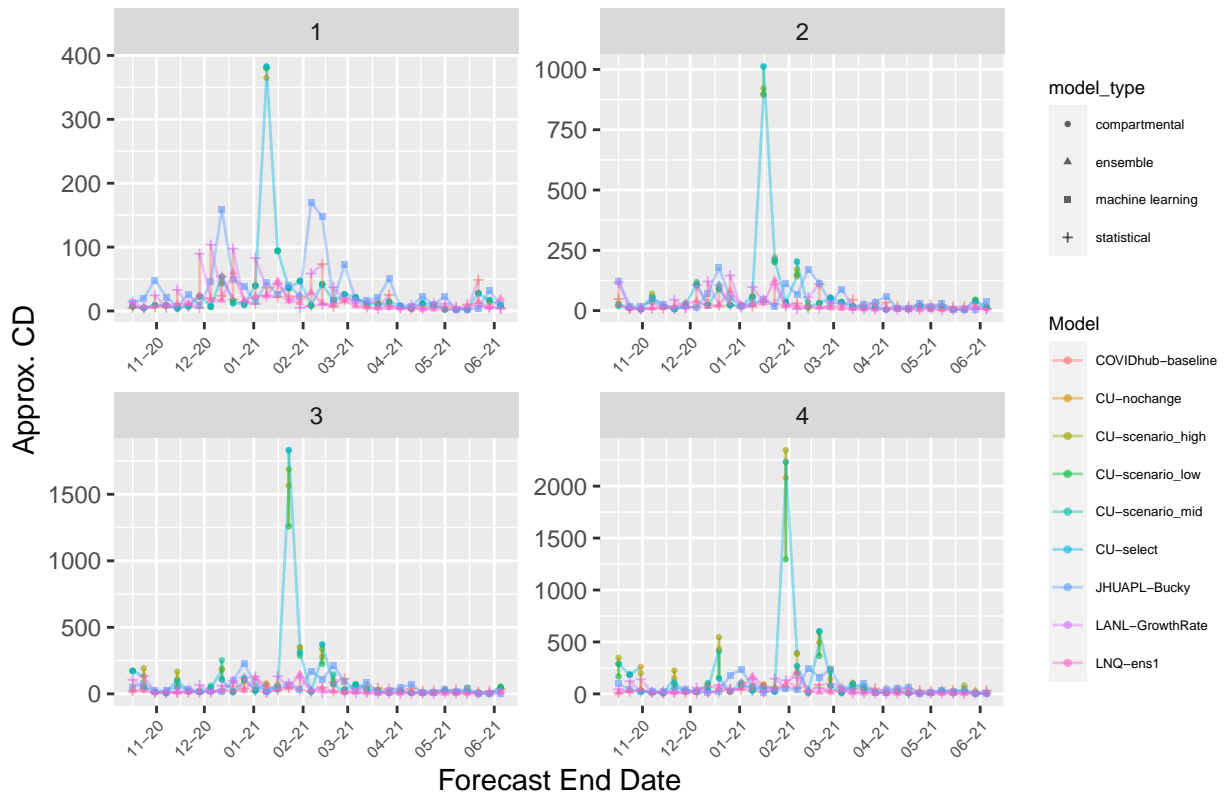
Below are the plots of approximated pairwise distances over time. Unlike indicent death targets, there are

too many model pairs to show (12 models for inc death and 17 models for inc case, so we have a combination of 6 choose 2 with no repeats (15) vs a combination of 11 choose 2 with no repeats (55)). So, I only pick the first 15 of the combinations:

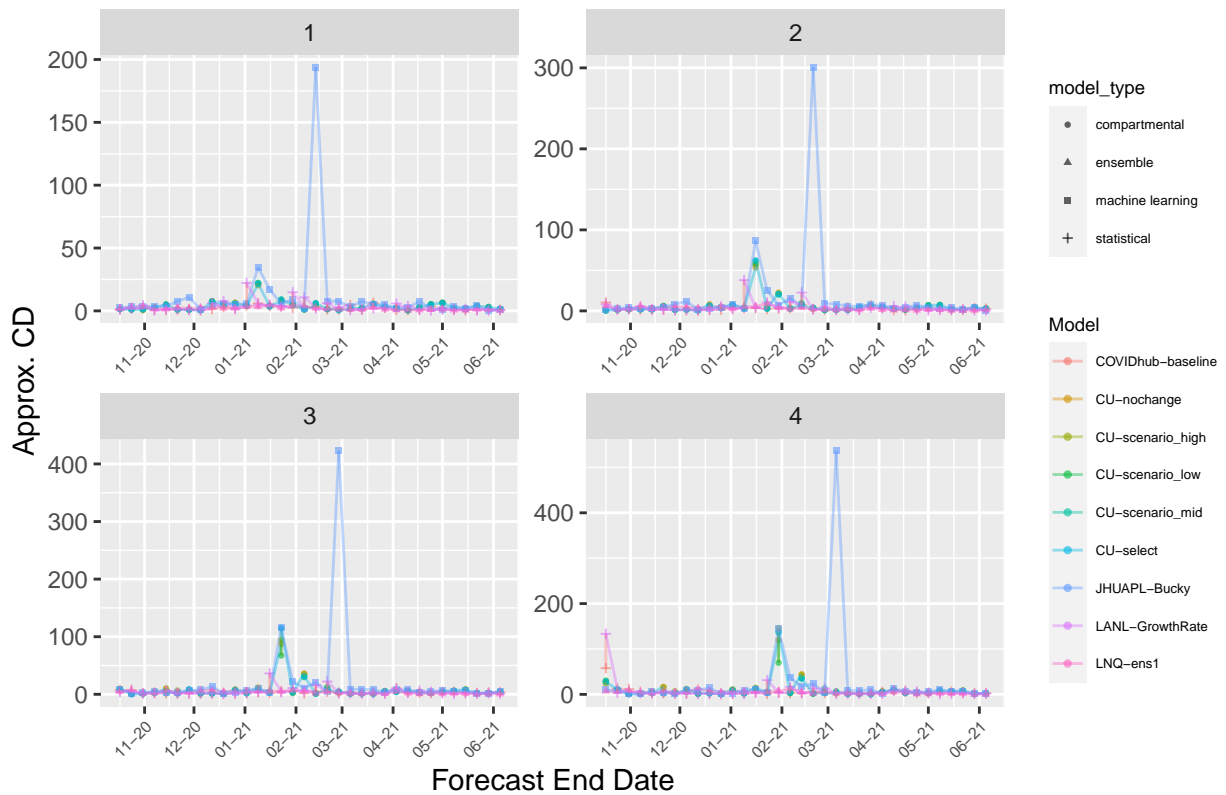
Approx. CD from COVIDhub–ensemble Over Time – AL



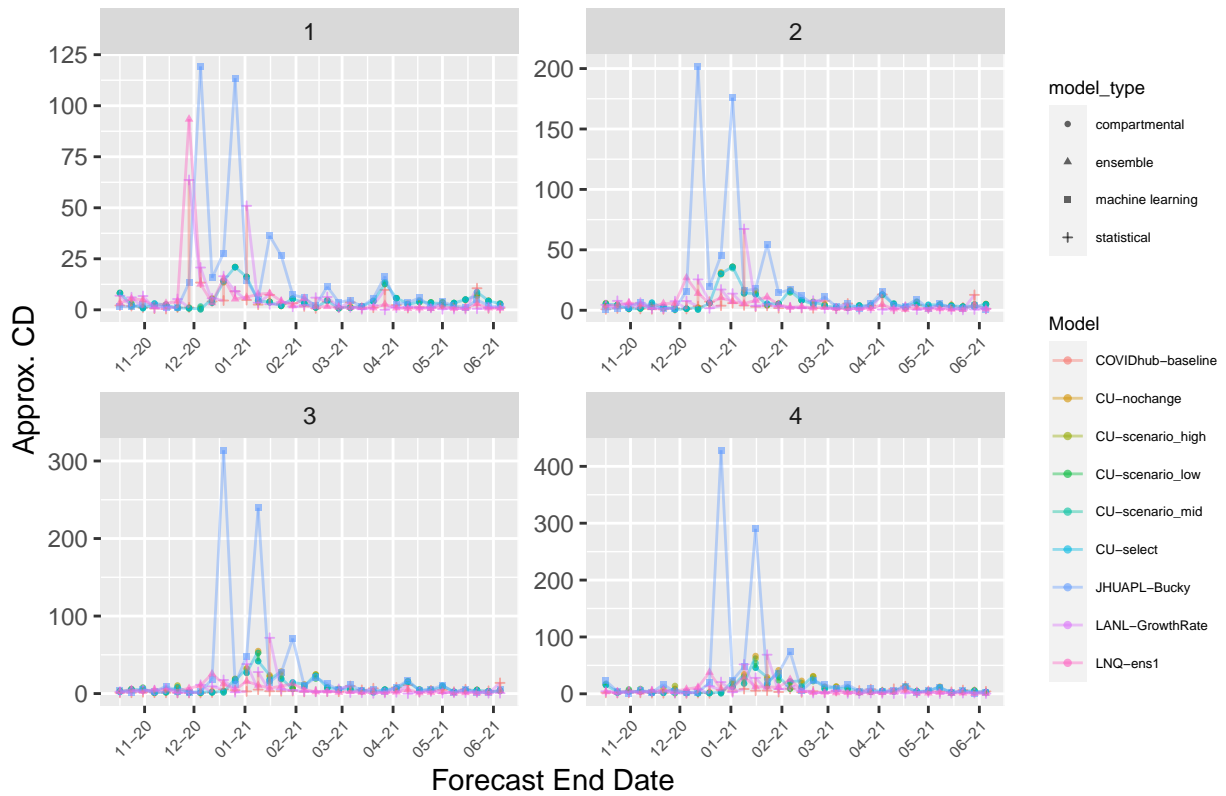
Approx. CD from COVIDhub–ensemble Over Time – AL



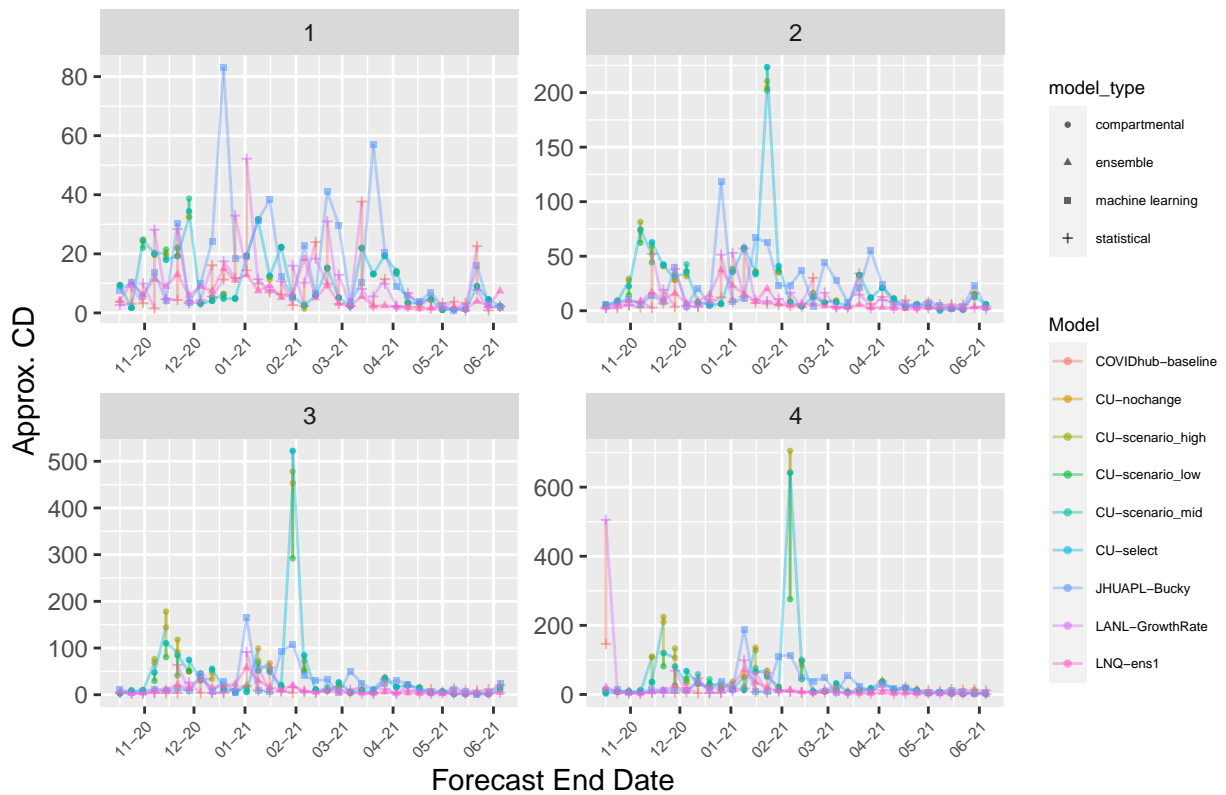
Approx. CD from COVIDhub–ensemble Over Time – AL

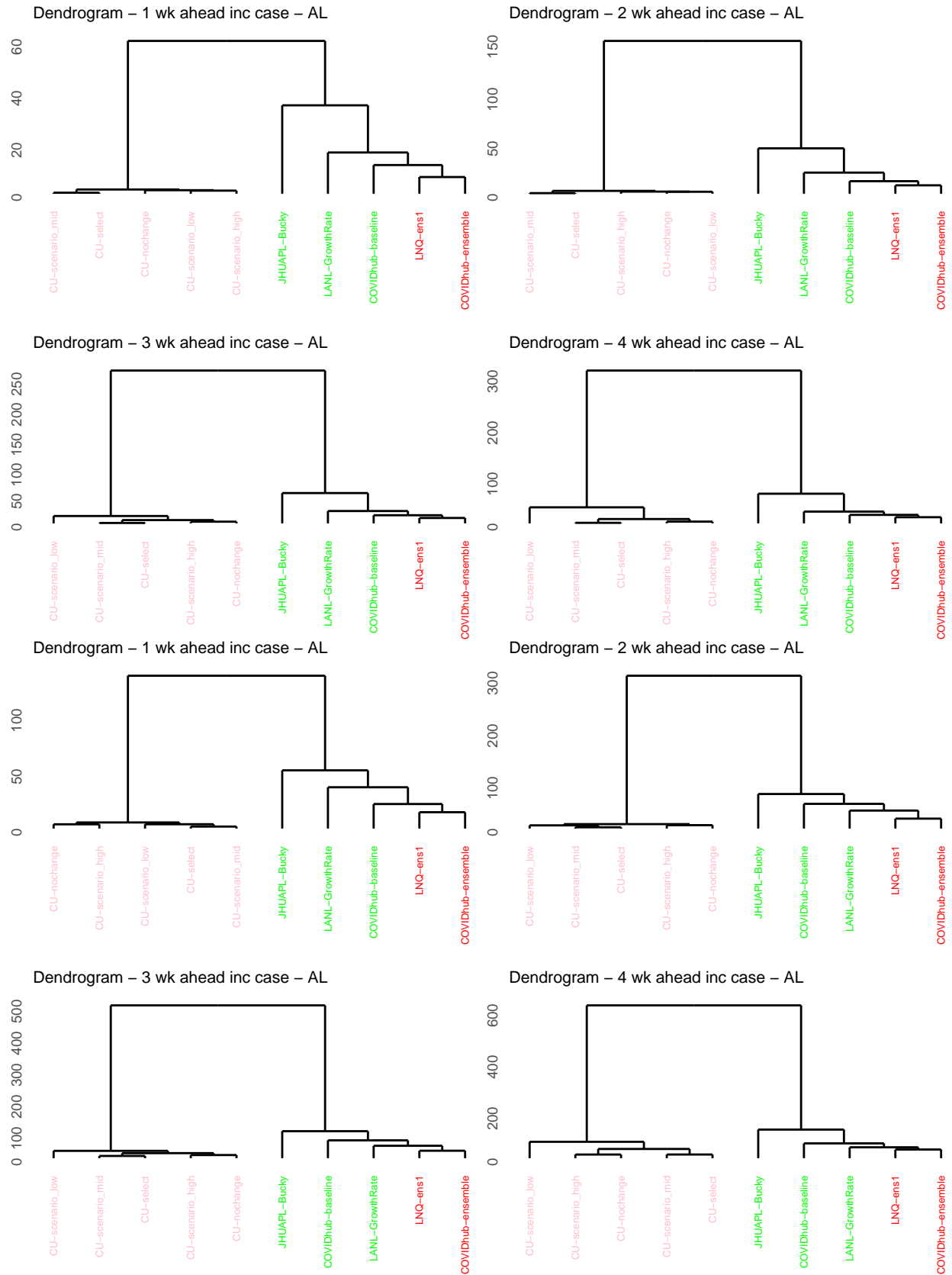


Approx. CD from COVIDhub-ensemble Over Time – AL

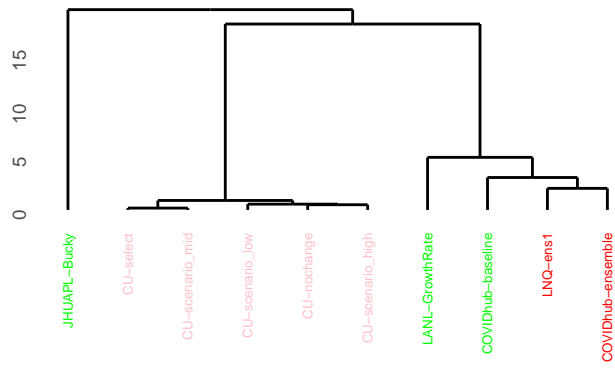


Approx. CD from COVIDhub-ensemble Over Time – AL

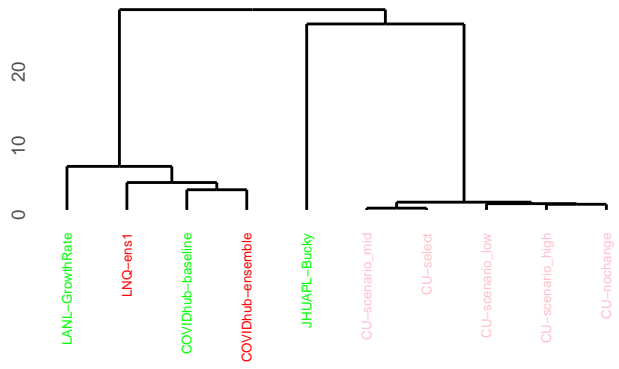




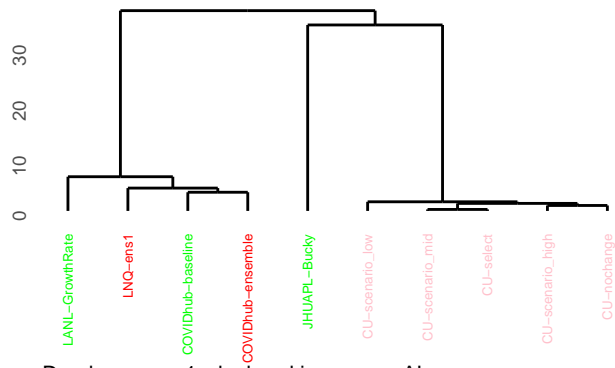
Dendrogram – 1 wk ahead inc case – AL



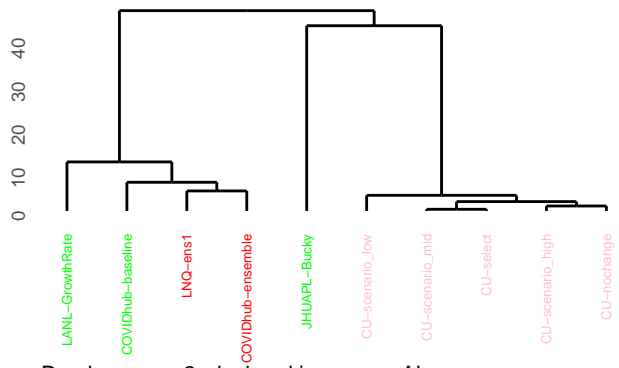
Dendrogram – 2 wk ahead inc case – AL



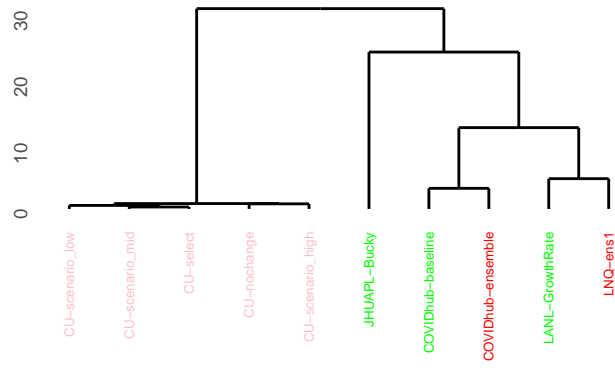
Dendrogram – 3 wk ahead inc case – AL



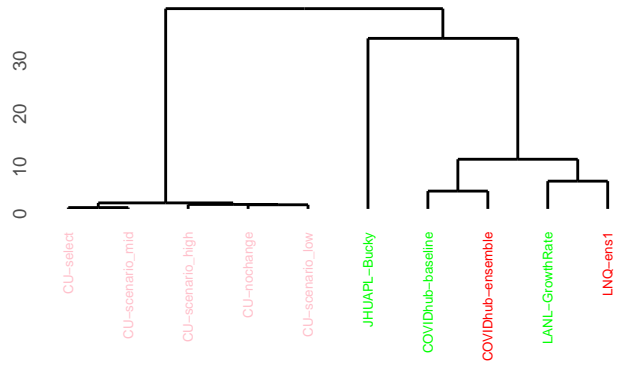
Dendrogram – 4 wk ahead inc case – AL



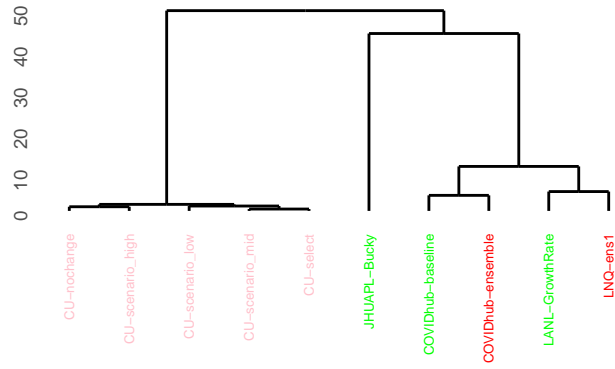
Dendrogram – 1 wk ahead inc case – AL



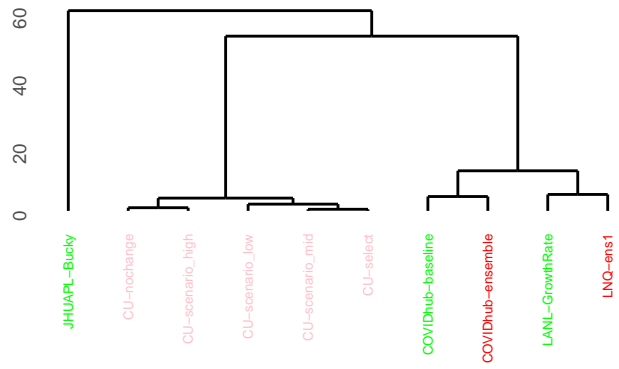
Dendrogram – 2 wk ahead inc case – AL



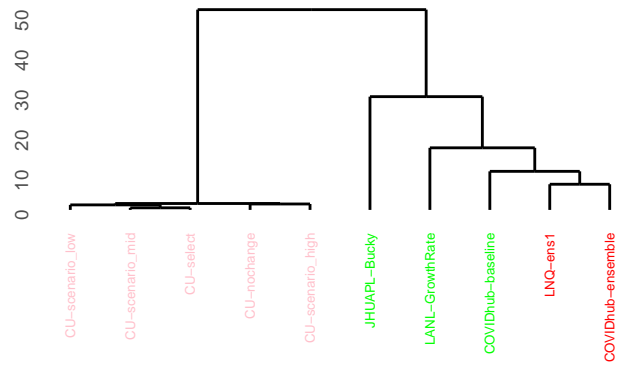
Dendrogram – 3 wk ahead inc case – AL



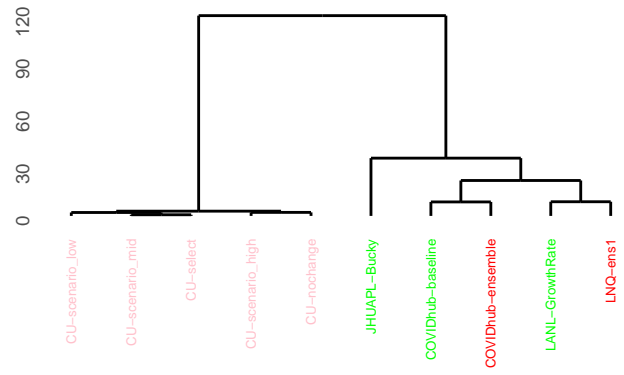
Dendrogram – 4 wk ahead inc case – AL



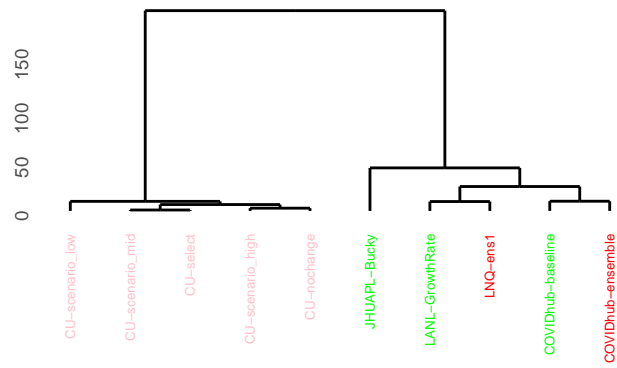
Dendrogram – 1 wk ahead inc case – AL



Dendrogram – 2 wk ahead inc case – AL



Dendrogram – 3 wk ahead inc case – AL



Dendrogram – 4 wk ahead inc case – AL

