Ensemble Comparison Project

Li Shandross

2023-08-31

Introduction/Background

Overview:

Goal: given quantile forecasts as an input, compare: - linear pools - direct computation of quantile average or median

Related papers:

- By Emily: https://royalsocietypublishing.org/doi/10.1098/rsif.2022.0659
- Busetti: https://onlinelibrary.wiley.com/doi/full/10.1111/obes.12163?casa_token=J9KIeOZWEUAAAAAA%3A8Xm3PyoD9TqAczQf1P6C9QOEfHA-KePb06cLGtlEdM-MhFpbrXIHmD1bdP9tO0gcGiwh4mjPlcXtNVA

Our setting:

- Primary interest is flu hospital admissions: will FluSight use the linear pool method this fall?
- But we only have 2 seasons of quantile forecasts for flu, and in the first season the data was iffy (full reporting only started halfway through the season). Therefore, if possible let's also include COVID hospital admissions

Methods

Start with flu:

- Both seasons
- All locations forecasted in the flusight exercise. Anticipating that there might be a difference, let's formally collect somewhere the locations used in flu so that we can easily use the same ones for covid if/when we add it in later.
- Include all forecast horizons, we can stratify results by horizon afterward
- All component models, no checks like whether the model provides all locations or all horizons. But the model should provide all quantile levels for any location/horizons where it provides any forecasts.

Ensemble methods, all equally weighted:

- Linear pool, using `hubEnsembles::linear_pool`
  - *Note:* we're calling distfromq to get from quantiles to full distributions. It makes some tail assumptions. The default is normal distributions for the tails. This is awkward because hospitalizations have to be non-negative, but a normal lower tail could extend into negative values. To address this, with this method let's truncate any negative quantile estimates to 0.
  - ELR is working on adding in a lognormal tail assumption to distfromq. When it is ready, we can add it into the comparison (so there would be 2 LP methods, LP-normal and LP-lognormal)
- Mean of quantiles (Vincent-mean), using `hubEnsembles::simple_ensemble`
- Median of quantiles (Vincent-median), using `hubEnsembles::simple_ensemble`

Data

Flu incident hospitalization truth and forecasts are queried from zoltar for the 2021-2022 and 2022-23 seasons. This is the only target/time period evaluated currently.

Forecast locations

Forecasts are made for all 50 states, Washington DC, Puerto Rico, the Virgin Islands come and the US as a whole. 1 - 4 week ahead incident flu hospitalizations Only validation phase, no testing phase (assume that will be the upcoming season).

Model specifications

Metrics and evaluation

- different ways to stratify evaluations: overall, by season, by horizon, by location, by (forecast) date
- separate by geographic scale (averaged states/territories vs US national) to avoid the US national results obscuring that of the state/territories
- Incident flu hospitalizations evaluated by usual horizon week, incident covid hospitalizations will also be evaluated on a weekly basis
- metrics: average wis, average mae, average 50% pi coverage, average 95% pi coverage, average rwis, average rmae

Results

Plot Forecasts

Will plot forecasts for all models, select locations (including US, highest count state, lowest count state/territory)

Overall model performance

Table 1: Summary of overall model performance across both seasons, averaged over the US national geographic scale.

model	wis	mae	cov50	cov95	rwis	rmae
median-ensemble	620.861	927.605	0.726	0.967	0.682	0.832
lp-normal	740.555	965.316	0.797	0.991	0.813	0.866
Flusight-baseline	910.836	1114.453	0.580	0.849	1.000	1.000
mean-ensemble	981.881	1064.878	0.693	0.962	1.078	0.956

Table 2: Summary of overall model performance across both seasons, averaged over the states geographic scale.

model	wis	mae	cov50	cov95	rwis	rmae
median-ensemble	18.158	27.360	0.597	0.922	0.794	0.933
lp-normal	19.745	27.932	0.709	0.990	0.863	0.953
mean-ensemble	20.180	29.582	0.595	0.889	0.882	1.009
Flusight-baseline	22.876	29.315	0.604	0.881	1.000	1.000

By season model performance

Table 3: Summary of by season model performance, averaged over the US national geographic scale.

model	season	wis	mae	cov50	cov95	rwis	rmae
median-ensemble	2021-2022	180.395	234.472	0.693	0.989	0.843	0.718
Flusight-baseline	2021-2022	213.998	326.591	0.625	1.000	1.000	1.000
mean-ensemble	2021-2022	214.566	292.673	0.636	1.000	1.003	0.896
lp-normal	2021-2022	253.189	267.105	0.773	1.000	1.183	0.818
median-ensemble	2022-2023	933.449	1419.505	0.750	0.952	0.664	0.848
lp-normal	2022-2023	1086.427	1460.821	0.815	0.984	0.773	0.873
Flusight-baseline	2022-2023	1405.366	1673.581	0.548	0.742	1.000	1.000
mean-ensemble	2022-2023	1526.427	1612.895	0.734	0.935	1.086	0.964

Table 4: Summary of by season model performance, averaged over the states geographic scale.

model	season	wis	mae	cov50	cov95	rwis	rmae
median-ensemble	2021-2022	8.038	12.194	0.570	0.926	0.845	0.908
mean-ensemble	2021-2022	8.398	12.540	0.551	0.896	0.882	0.934
lp-normal	2021-2022	9.074	12.466	0.691	0.993	0.953	0.928
Flusight-baseline	2021-2022	9.518	13.433	0.650	0.922	1.000	1.000
median-ensemble	2022-2023	25.341	38.124	0.617	0.919	0.783	0.939
lp-normal	2022-2023	27.317	38.909	0.721	0.988	0.844	0.959
mean-ensemble	2022-2023	28.541	41.676	0.626	0.885	0.882	1.027
Flusight-baseline	2022-2023	32.356	40.586	0.572	0.851	1.000	1.000

Model performance by horizon week

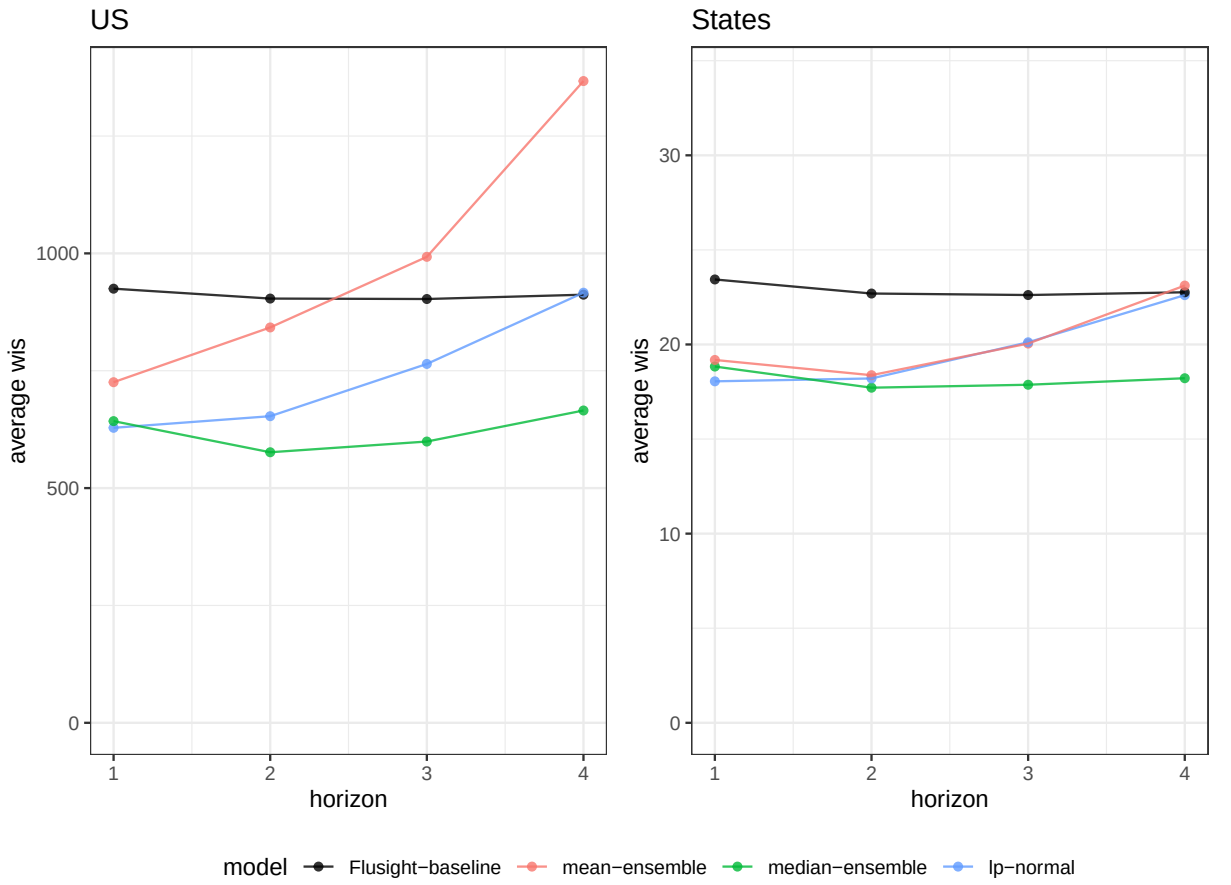
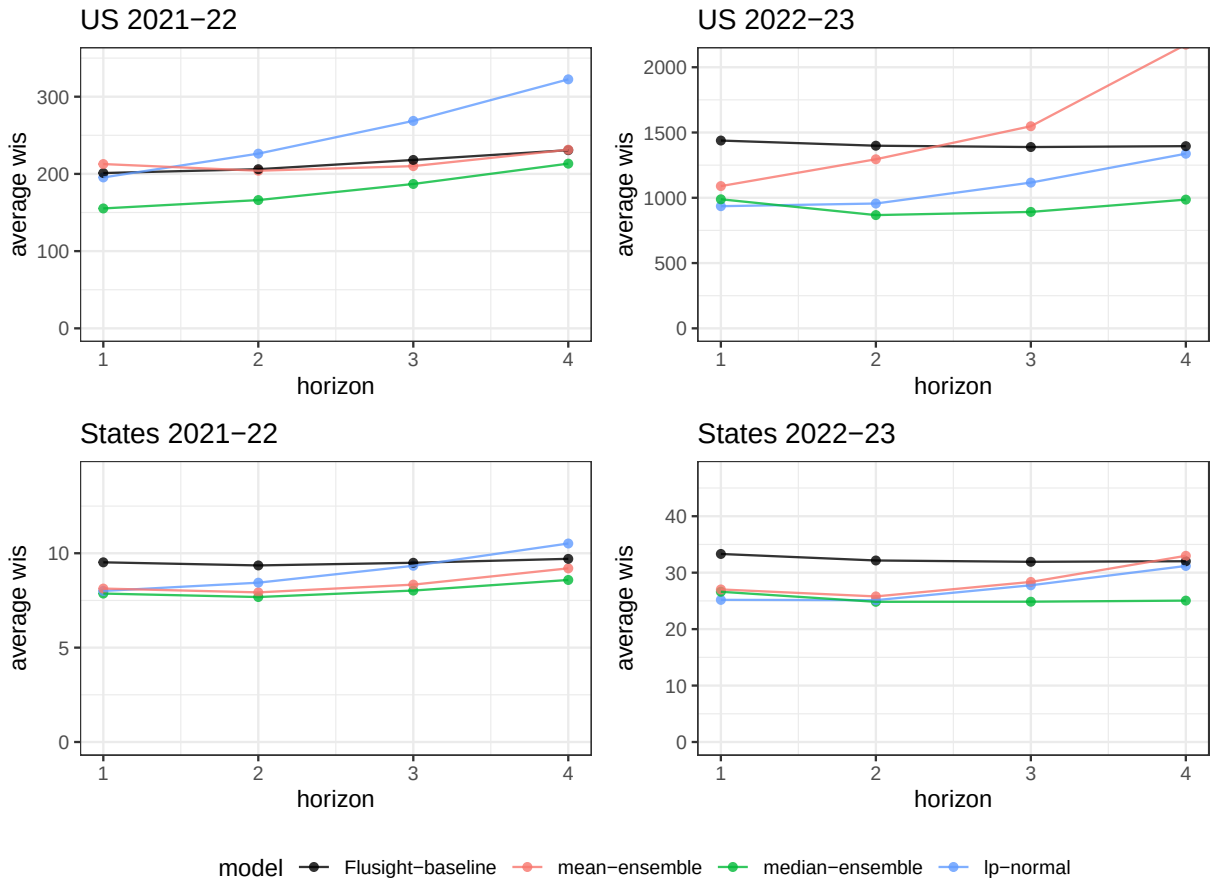


Figure 1: Average WIS by the horizon week for each model for the US national level and the states level.

Additional by season split



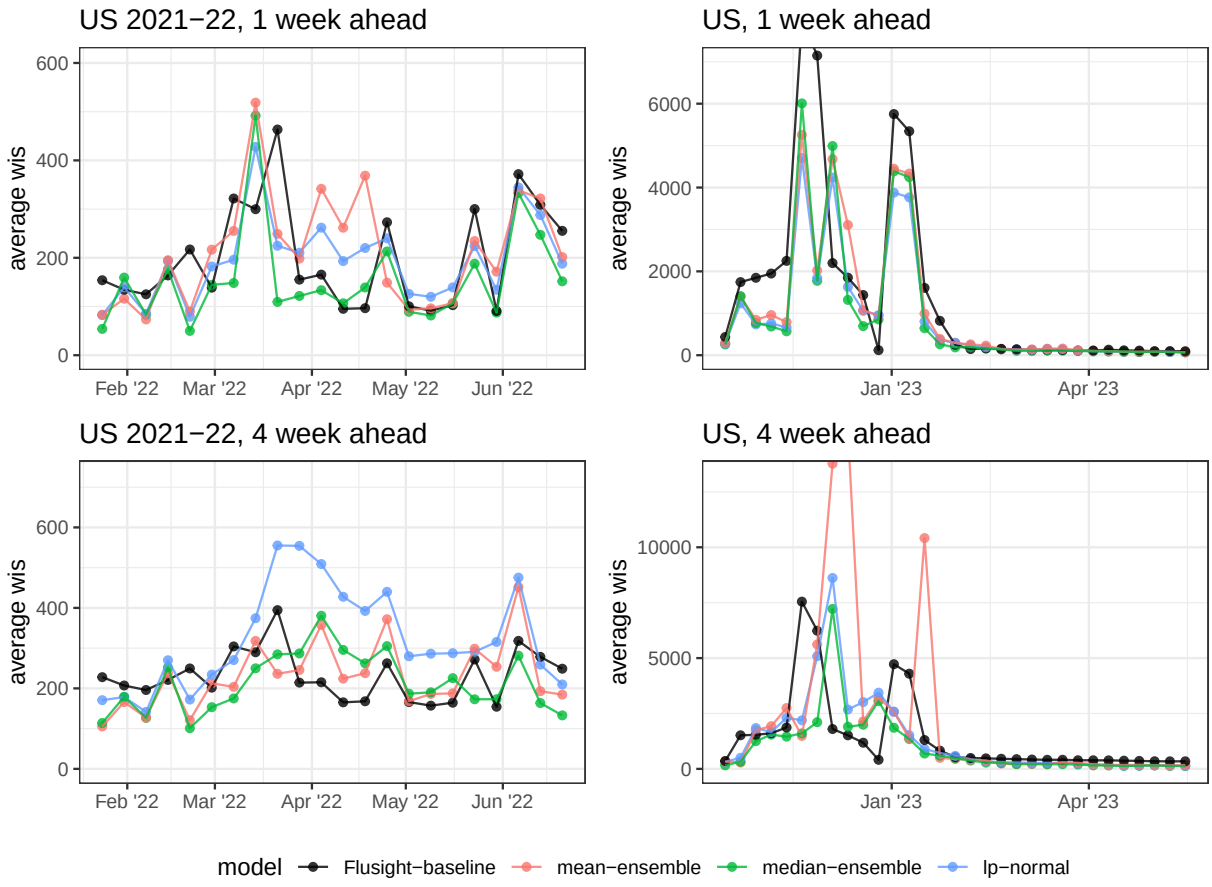


Figure 2: Average h-week ahead WIS and 95% PI coverage for each model for the US national level.

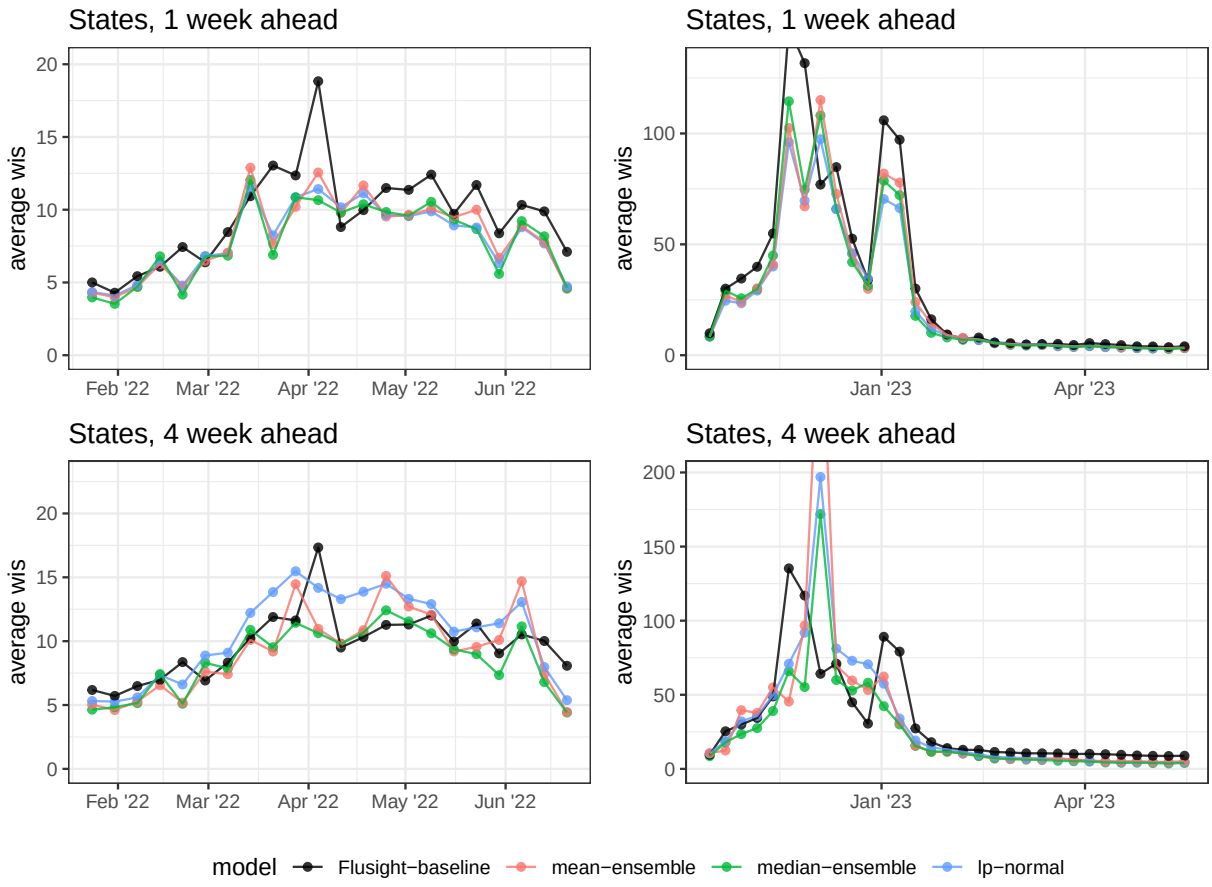
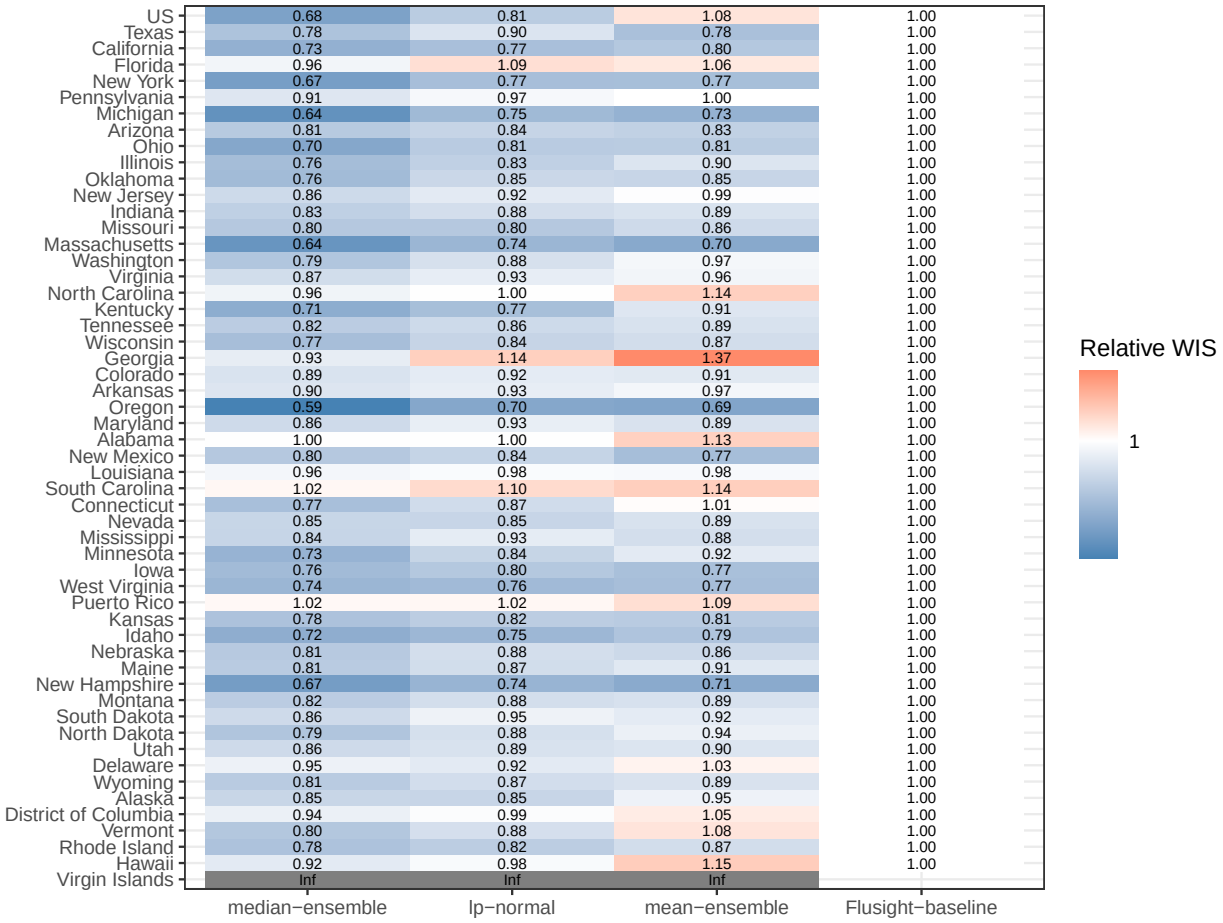
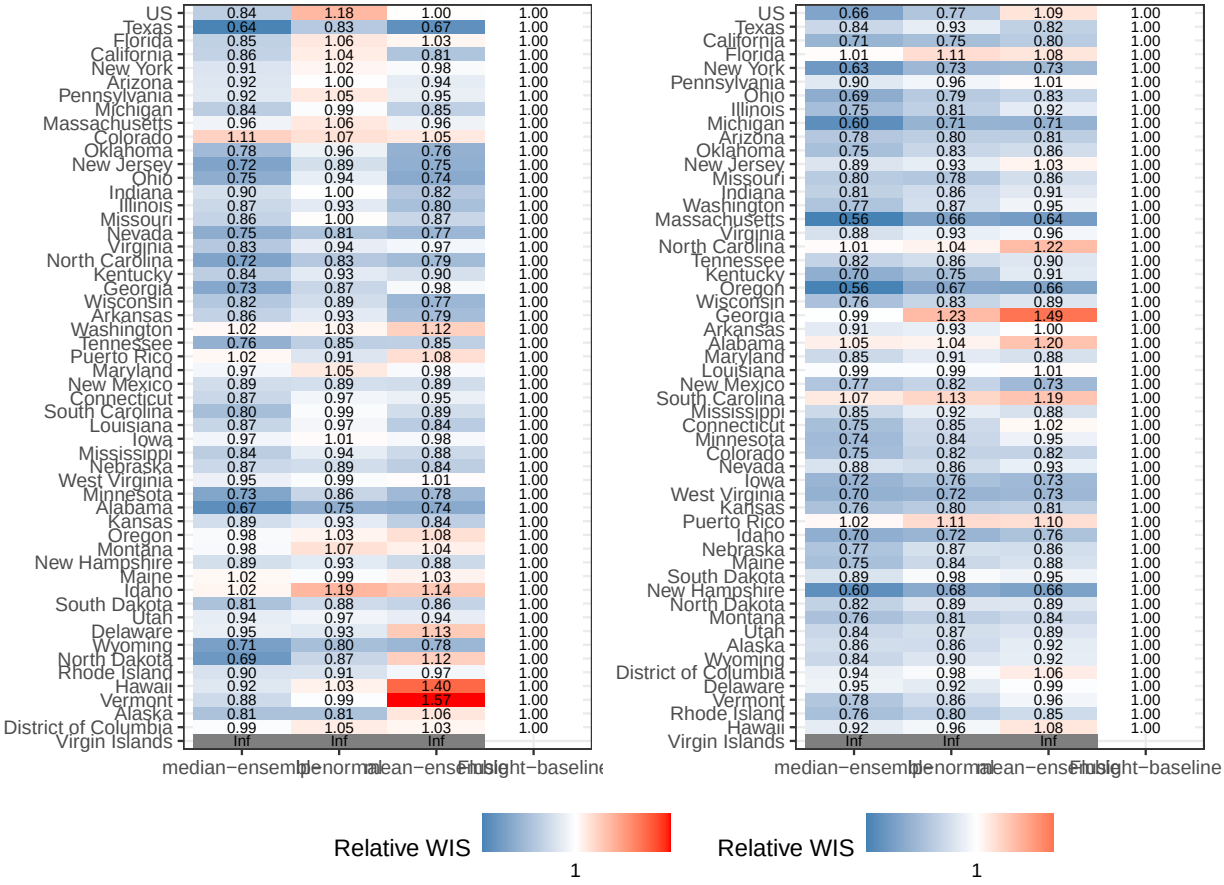


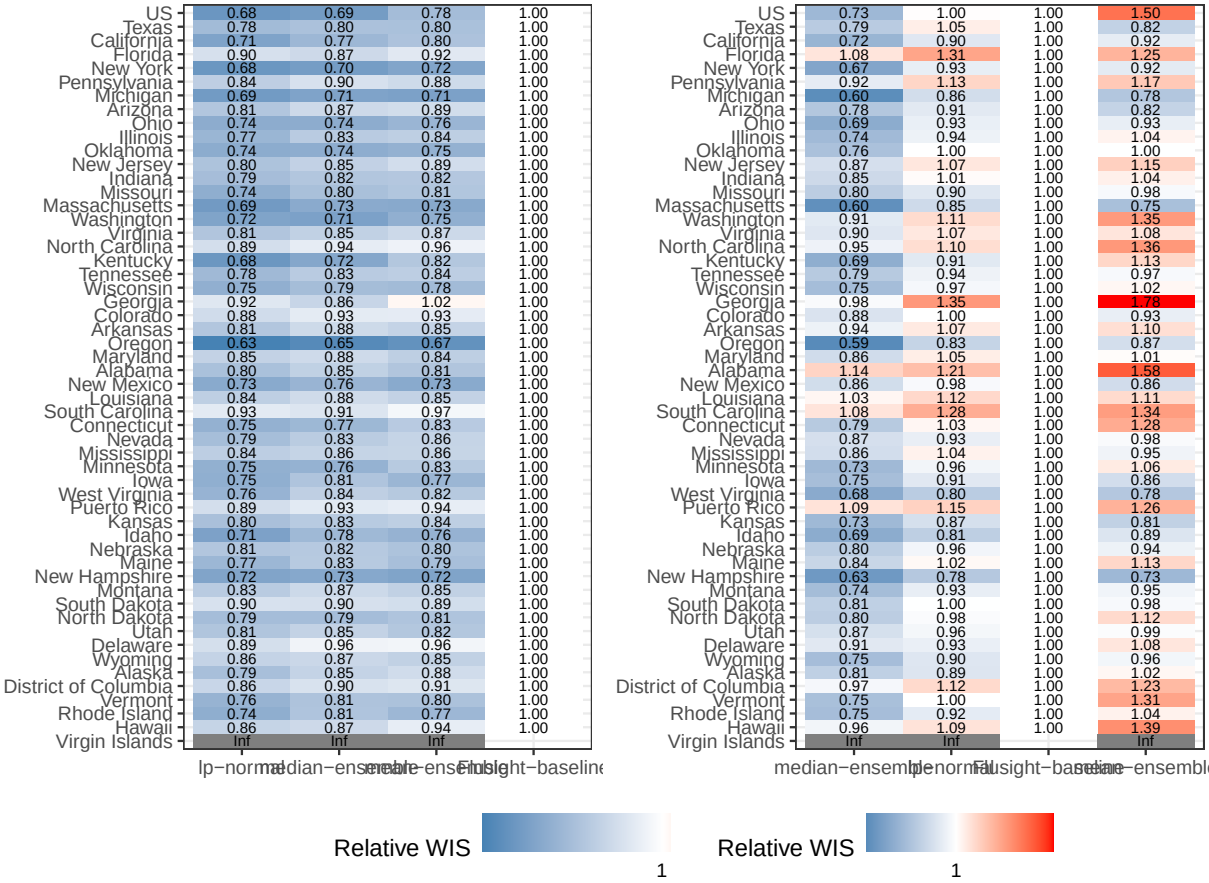
Figure 3: Average h-week ahead WIS and 95% PI coverage for each model for the States national level.

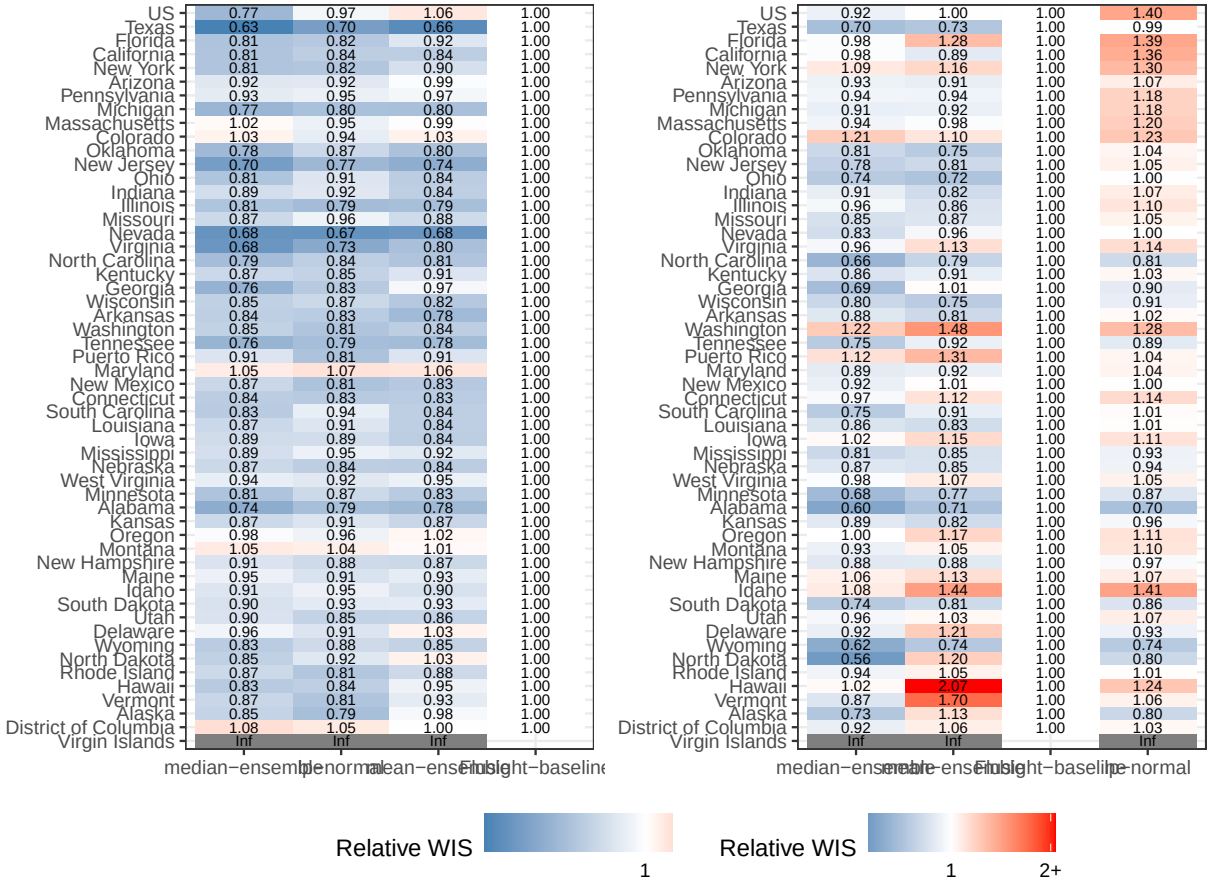
Model Performance by Week

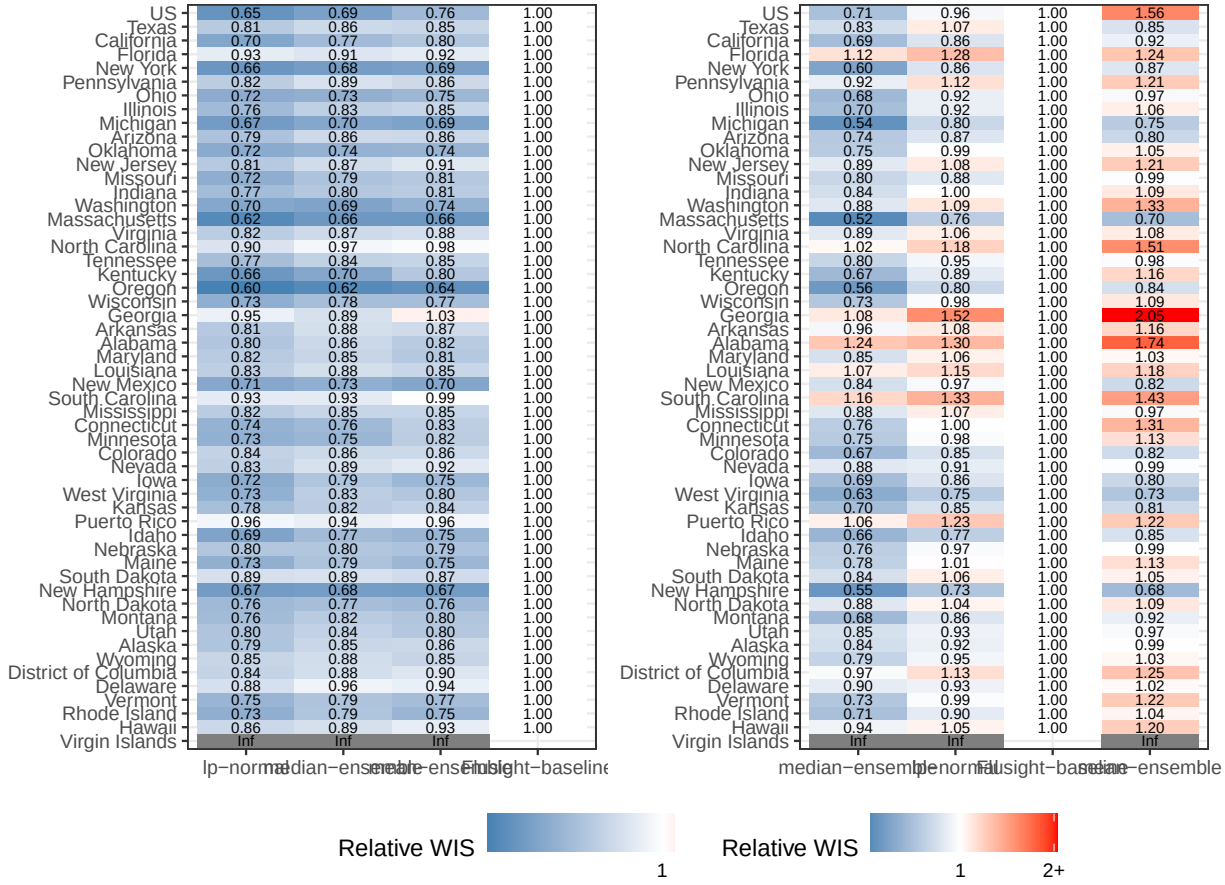
Model performance by location




US 0.84 1.18 1.00 1.00
Texas 0.64 0.83 0.67 1.00
Florida 0.85 1.06 1.03 1.00
California 0.86 1.04 0.81 1.00
New York 0.91 1.02 0.98 1.00
Arizona 0.92 1.00 0.94 1.00
Pennsylvania 0.92 1.05 0.95 1.00
Michigan 0.84 0.99 0.85 1.00
Massachusetts 0.96 1.06 0.96 1.00
Colorado 1.11 1.07 1.05 1.00
Oklahoma 0.78 0.96 0.76 1.00
New Jersey 0.72 0.89 0.75 1.00
Ohio 0.75 0.94 0.74 1.00
Indiana 0.90 1.00 0.82 1.00
Illinois 0.87 0.93 0.80 1.00
Missouri 0.86 1.00 0.87 1.00
Nevada 0.75 0.81 0.77 1.00
Virginia 0.83 0.94 0.97 1.00
North Carolina 0.72 0.83 0.79 1.00
Kentucky 0.84 0.93 0.90 1.00
Georgia 0.73 0.87 0.98 1.00
Wisconsin 0.82 0.89 0.77 1.00
Arkansas 0.86 0.93 0.79 1.00
Washington 1.02 1.03 1.12 1.00
Tennessee 0.76 0.85 0.85 1.00
Puerto Rico 1.02 0.91 1.08 1.00
Maryland 0.97 1.05 0.98 1.00
New Mexico 0.89 0.89 0.89 1.00
Connecticut 0.87 0.97 0.95 1.00
South Carolina 0.80 0.99 0.89 1.00
Louisiana 0.87 0.97 0.84 1.00
Iowa 0.97 1.01 0.98 1.00
Mississippi 0.84 0.94 0.88 1.00
Nebraska 0.87 0.89 0.84 1.00
West Virginia 0.95 0.99 1.01 1.00
Minnesota 0.73 0.86 0.78 1.00
Alabama 0.67 0.75 0.74 1.00
Kansas 0.89 0.93 0.84 1.00
Oregon 0.98 1.03 1.08 1.00
Montana 0.98 1.07 1.04 1.00
New Hampshire 0.89 0.93 0.88 1.00
Maine 1.02 0.99 1.03 1.00
Idaho 1.02 1.19 1.14 1.00
South Dakota 0.81 0.88 0.86 1.00
Utah 0.94 0.97 0.94 1.00
Delaware 0.95 0.93 1.13 1.00
Wyoming 0.71 0.80 0.78 1.00
North Dakota 0.69 0.87 1.12 1.00
Rhode Island 0.90 0.91 0.97 1.00
Hawaii 0.92 1.03 1.40 1.00
Vermont 0.88 0.99 1.57 1.00
Alaska 0.81 0.81 1.06 1.00
District of Columbia 0.99 1.05 1.03 1.00
Virgin Islands Inf Inf Inf
median−ensemble
plenormal
mean−ensemble
Flusight−baseline
Relative WIS
1
US 0.66 0.77 1.09 1.00
Texas 0.84 0.93 0.82 1.00
California 0.71 0.75 0.80 1.00
Florida 1.01 1.11 1.08 1.00
New York 0.63 0.73 0.73 1.00
Pennsylvania 0.90 0.96 1.01 1.00
Ohio 0.69 0.79 0.83 1.00
Illinois 0.75 0.81 0.92 1.00
Michigan 0.60 0.71 0.71 1.00
Arizona 0.78 0.80 0.81 1.00
Oklahoma 0.75 0.83 0.86 1.00
New Jersey 0.89 0.93 1.03 1.00
Missouri 0.80 0.78 0.86 1.00
Indiana 0.81 0.86 0.91 1.00
Washington 0.77 0.87 0.95 1.00
Massachusetts 0.56 0.66 0.64 1.00
Virginia 0.88 0.93 0.96 1.00
North Carolina 1.01 1.04 1.22 1.00
Tennessee 0.82 0.86 0.90 1.00
Kentucky 0.70 0.75 0.91 1.00
Oregon 0.56 0.67 0.66 1.00
Wisconsin 0.76 0.83 0.89 1.00
Georgia 0.99 1.23 1.49 1.00
Arkansas 0.91 0.93 1.00 1.00
Alabama 1.05 1.04 1.20 1.00
Maryland 0.85 0.91 0.88 1.00
Louisiana 0.99 0.99 1.01 1.00
New Mexico 0.77 0.82 0.73 1.00
South Carolina 1.07 1.13 1.19 1.00
Mississippi 0.85 0.92 0.88 1.00
Connecticut 0.75 0.85 1.02 1.00
Minnesota 0.74 0.84 0.95 1.00
Colorado 0.75 0.82 0.82 1.00
Nevada 0.88 0.86 0.93 1.00
Iowa 0.72 0.76 0.73 1.00
West Virginia 0.70 0.72 0.73 1.00
Kansas 0.76 0.80 0.81 1.00
Puerto Rico 1.02 1.11 1.10 1.00
Idaho 0.70 0.72 0.76 1.00
Nebraska 0.77 0.87 0.86 1.00
Maine 0.75 0.84 0.88 1.00
South Dakota 0.89 0.98 0.95 1.00
New Hampshire 0.60 0.68 0.66 1.00
North Dakota 0.82 0.89 0.89 1.00
Montana 0.76 0.81 0.84 1.00
Utah 0.84 0.87 0.89 1.00
Alaska 0.86 0.86 0.92 1.00
Wyoming 0.84 0.90 0.92 1.00
District of Columbia 0.94 0.98 1.06 1.00
Delaware 0.95 0.92 0.99 1.00
Vermont 0.78 0.86 0.96 1.00
Rhode Island 0.76 0.80 0.85 1.00
Hawaii 0.92 0.96 1.08 1.00
Virgin Islands Inf Inf Inf
median−ensemble
plenormal
mean−ensemble
Flusight−baseline
Relative WIS
1


US
Texas
California
Florida
New York
Pennsylvania
Michigan
Arizona
Ohio
Illinois
Oklahoma
New Jersey
Indiana
Missouri
Massachusetts
Washington
Virginia
North Carolina
Kentucky
Tennessee
Wisconsin
Georgia
Colorado
Arkansas
Oregon
Maryland
Alabama
New Mexico
Louisiana
South Carolina
Connecticut
Nevada
Mississippi
Minnesota
Iowa
West Virginia
Puerto Rico
Kansas
Idaho
Nebraska
Maine
New Hampshire
Montana
South Dakota
North Dakota
Utah
Delaware
Wyoming
Alaska
District of Columbia
Vermont
Rhode Island
Hawaii
Virgin Islands
Relative WIS
1
Relative WIS
1


US
Texas
Florida
California
New York
Arizona
Pennsylvania
Michigan
Massachusetts
Colorado
Oklahoma
New Jersey
Ohio
Indiana
Illinois
Missouri
Nevada
Virginia
North Carolina
Kentucky
Georgia
Wisconsin
Arkansas
Washington
Tennessee
Puerto Rico
Maryland
New Mexico
Connecticut
South Carolina
Louisiana
Iowa
Mississippi
Nebraska
West Virginia
Minnesota
Alabama
Kansas
Oregon
Montana
New Hampshire
Maine
Idaho
South Dakota
Utah
Delaware
Wyoming
North Dakota
Rhode Island
Hawaii
Vermont
Alaska
District of Columbia
Virgin Islands
median−ensemble
lp−normal
mean−ensemble
Flusight−baseline
Relative WIS
1
2+



Conclusions/Discussion

References