Infectious Disease Forecast Evaluation Via Social Utility: Allocation Scores

MIDAS 2023

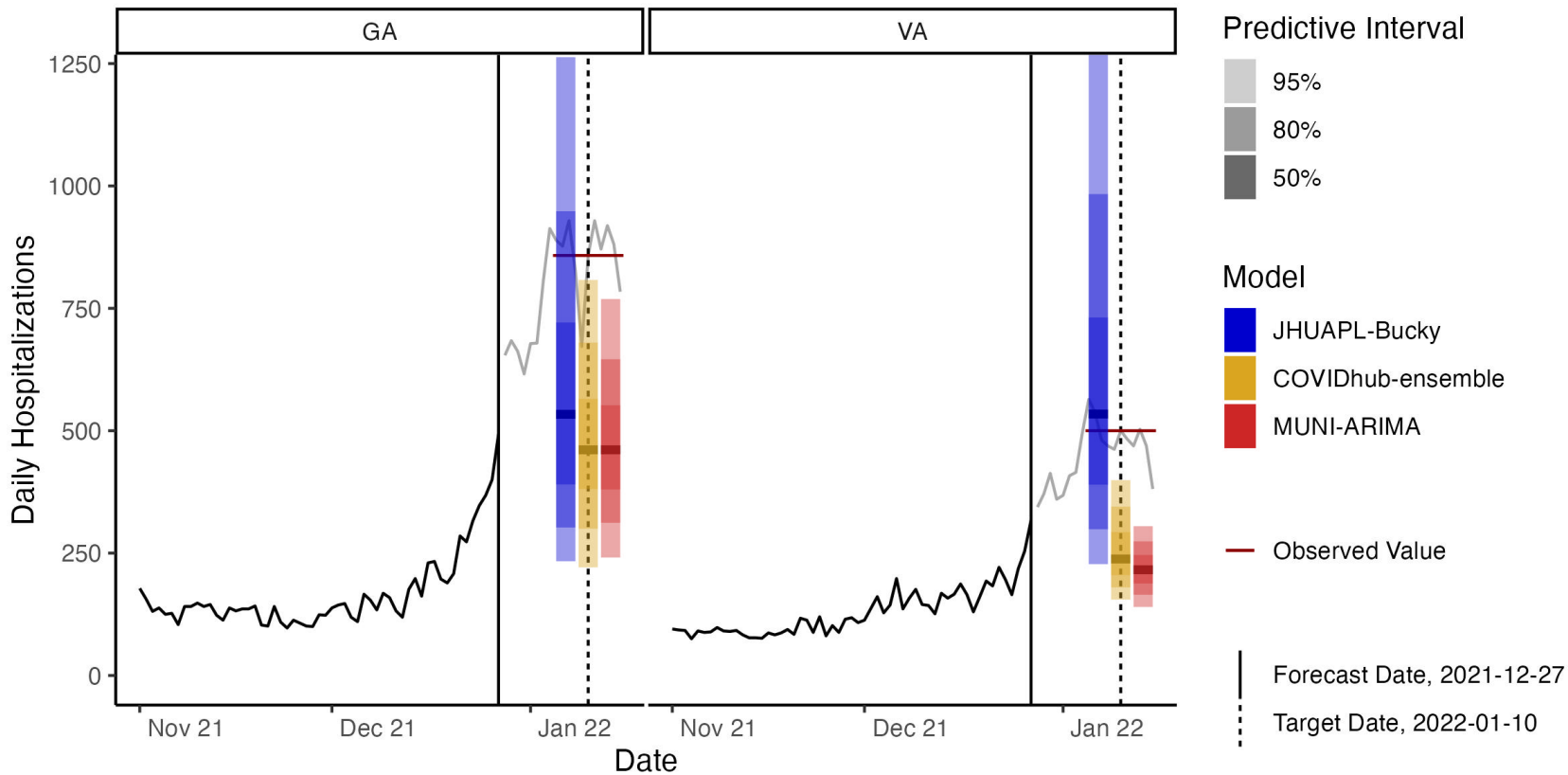
Aaron Gerding, Nick Reich, Ben Rogers, Evan Ray

School of Public Health and Health Sciences, UMass, Amherst
Department of Biostatistics and Epidemiology

Outbreak Forecast Hubs and **Informed** Resource Allocation

Hubs developed to **"inform public health responses"**

- such as how resources are **allocated** among locations
- e.g., medical supplies, facility capacity, personnel



Scoring rules and social welfare

Hubs strive to rank and combine forecasts so as to optimize social welfare via the public health decisions that forecasts inform

- uncertainty quantification, and therefore, probabilistic forecasting essential
- basic strategy: use a **scoring rule** S which assigns a loss $S(F, y)$ when a probabilistic forecast F of Y is chosen and $Y = y$ is observed.
- optimizing welfare requires that forecasters say what they believe; so S should be **proper** meaning $E_F[S(F, Y)] \leq E_F S(G, Y)]$ for all F, G

Current standard is the **Weighted Interval Score** (discrete CRPS)

- adopted largely for convenient scoring of quantile forecasts.

Tools from decision theory

A central goal in design of scoring rules

Tie forecast scores directly to the benefits to society of the decisions they inform

Key tools from decision theory for linking success/failure of forecast-informed policy actions to scoring rules:

- Let $l(x, y)$ be the **loss** of experiencing y after taking policy action x
- The **Bayes risk** of a forecast $Y \sim F$ is $\min_x E_F[l(x, Y)]$
- A **Bayes act** for F is an action x^F that attains this minimum
- Losses from Bayes acts define an *automatically proper* scoring rule

$$S(F, y) := l(x^F, y)$$

Note: Scoring rules are really only a means to an end. It is the expected score, in this case the Bayes risk, which characterize the value a forecast adds for a decision maker. Scoring rule sample averages estimate the expected score.

A new scoring rule via a new loss function for constrained actions

Our basic example:

- x and y are vectors in $\mathbb{N}_0^{52}$
- $y =$ number of severe cases in US states and territories
- $l(x, y) =$ unmet need when x beds allocated and y severe cases occur

$$l(x, y) = \sum_{i=1}^{52} \max(y_i - x_i, 0)$$

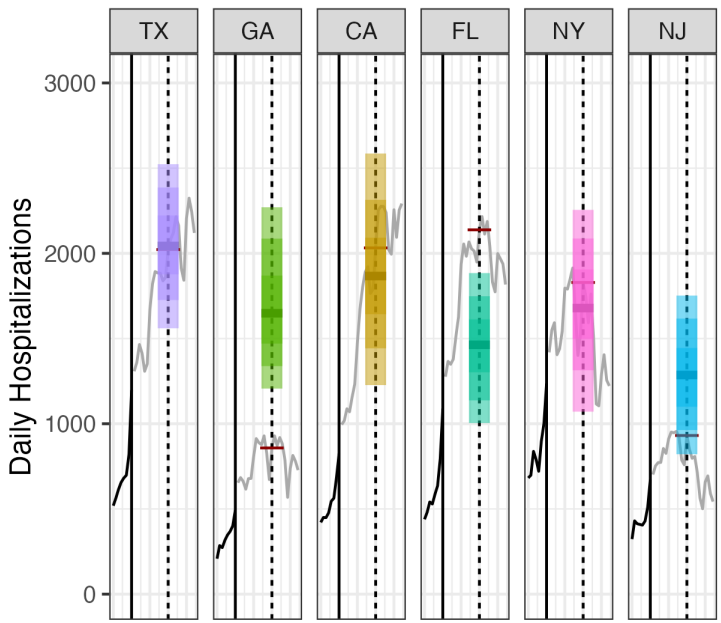
$$x^F = \text{minimizer of } E_F[l(x, Y)] \text{ over all feasible } x$$

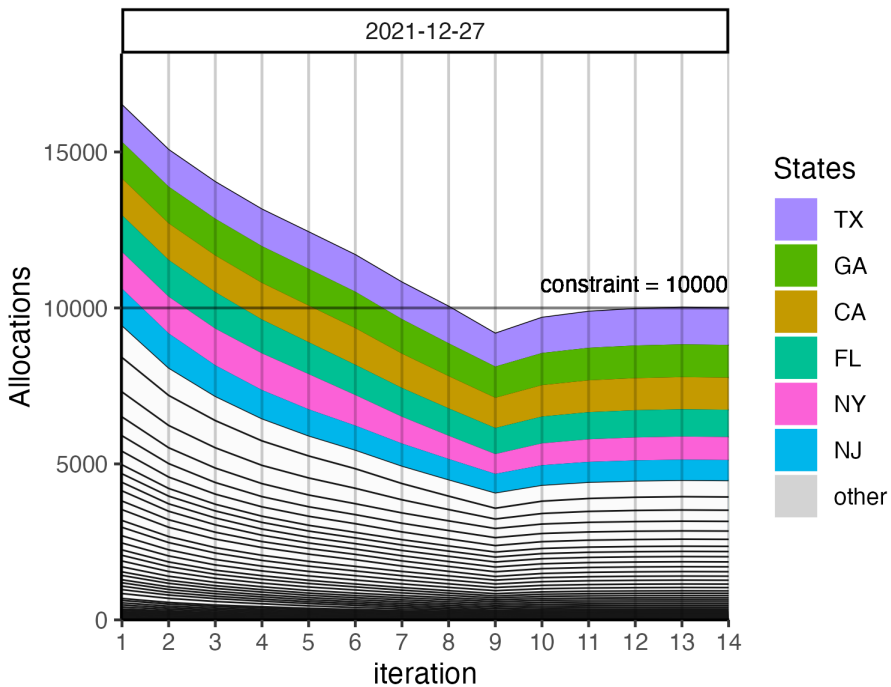
Our new idea: Define **feasible** x as satisfying $\sum x_i \leq K$

Computation

Obtaining x^F is a constrained stochastic optimization problem

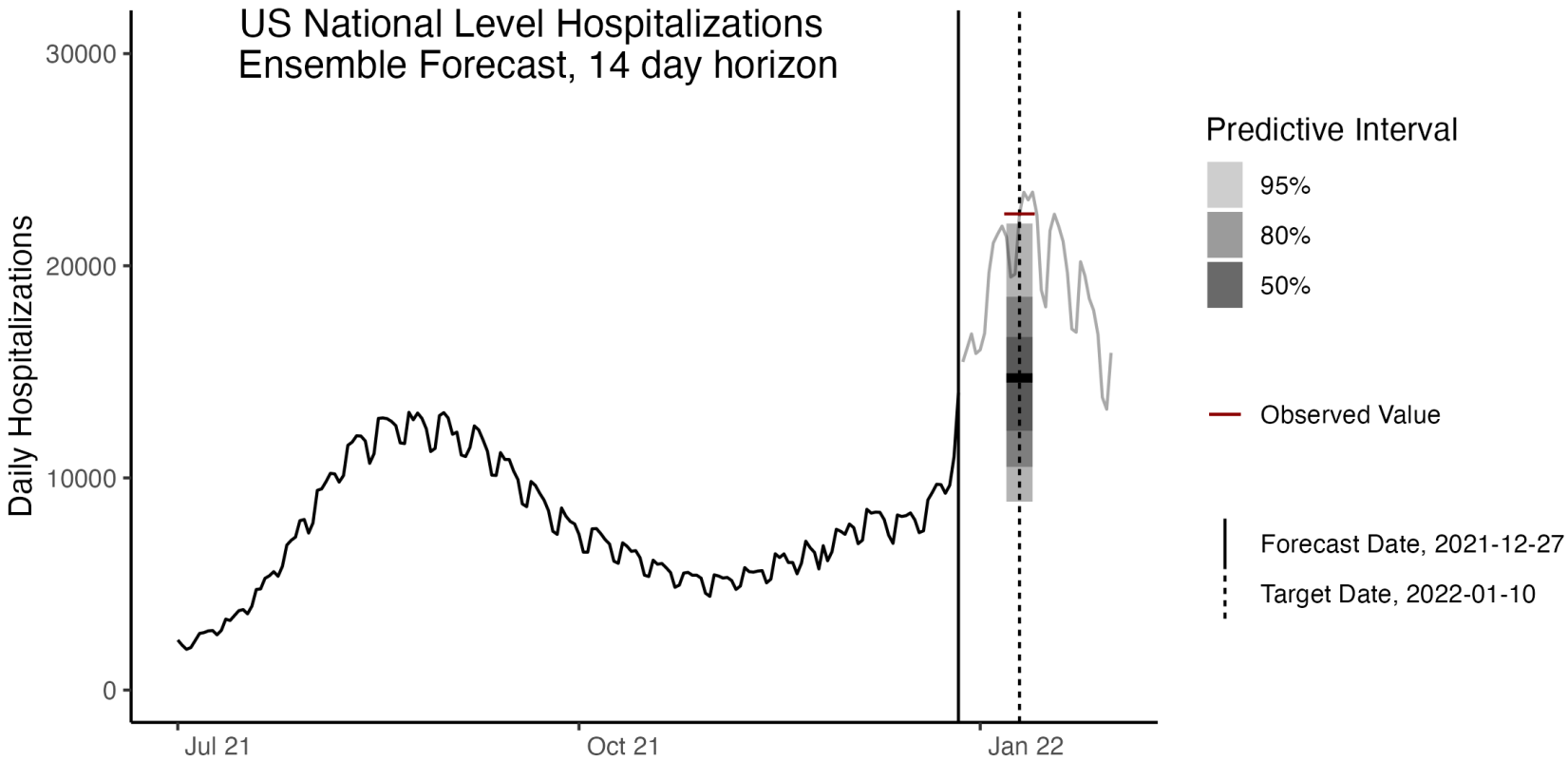
- known in inventory management as a **constrained multi-product newsvendor** problem
- formally solvable using Lagrange multiplier method to get a quantile representation $x_i^F = F_i^{-1}(\tau(K, F)), i = 1, \ldots, N$
- in practice, we find x_i^F's via an iterative method:



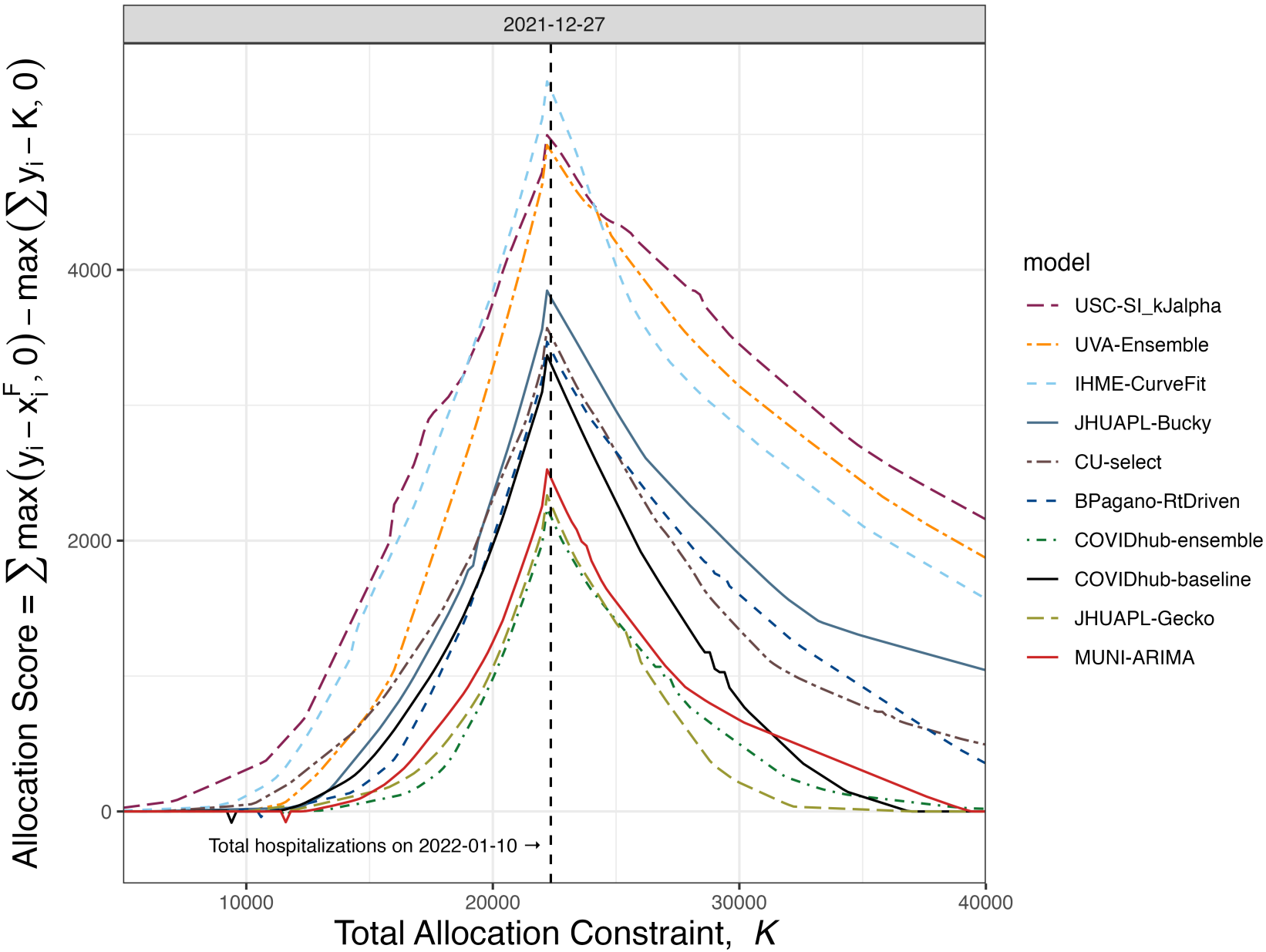


Application

December 2021: Omicron wave clearly started US but forecast teams unsure of severity given uncertainty about R_0, cross-protection by vaccination, previous infection, etc.

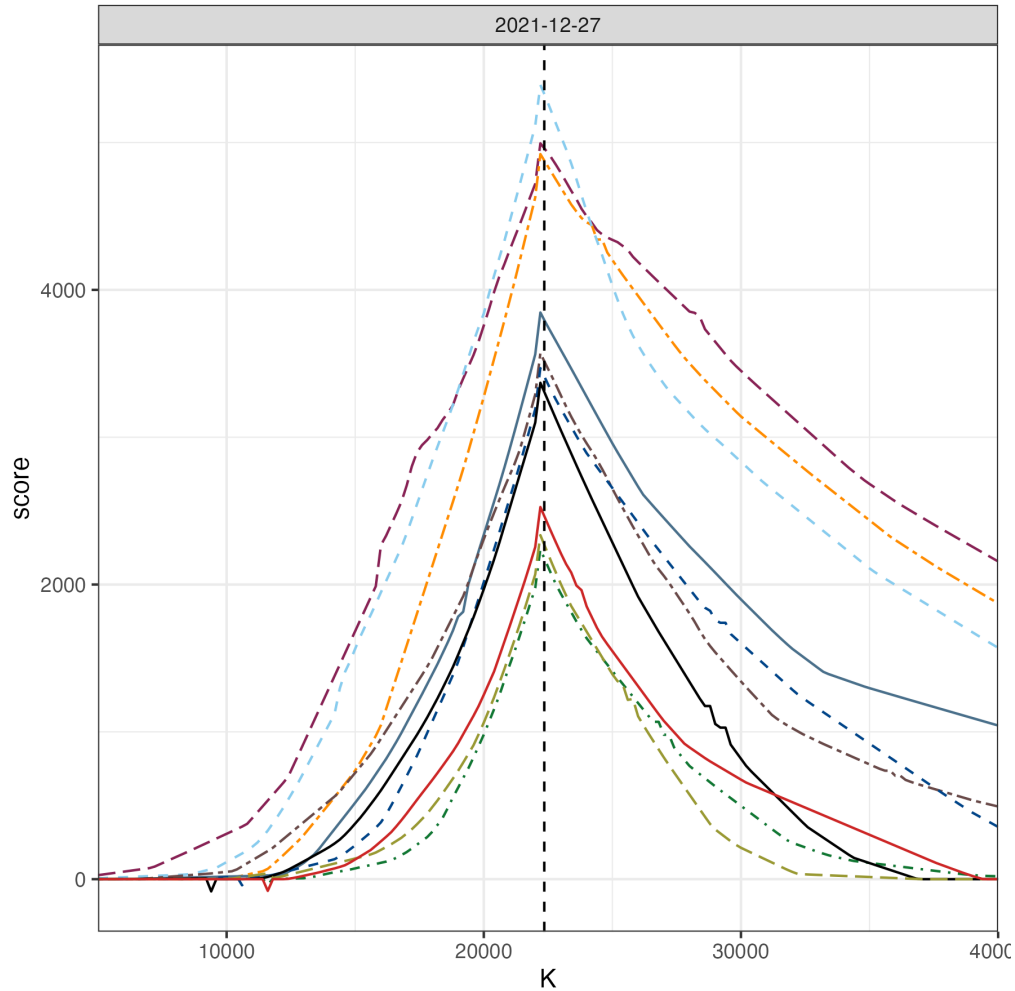


Oracle adjusted allocation scores near Omicron peak

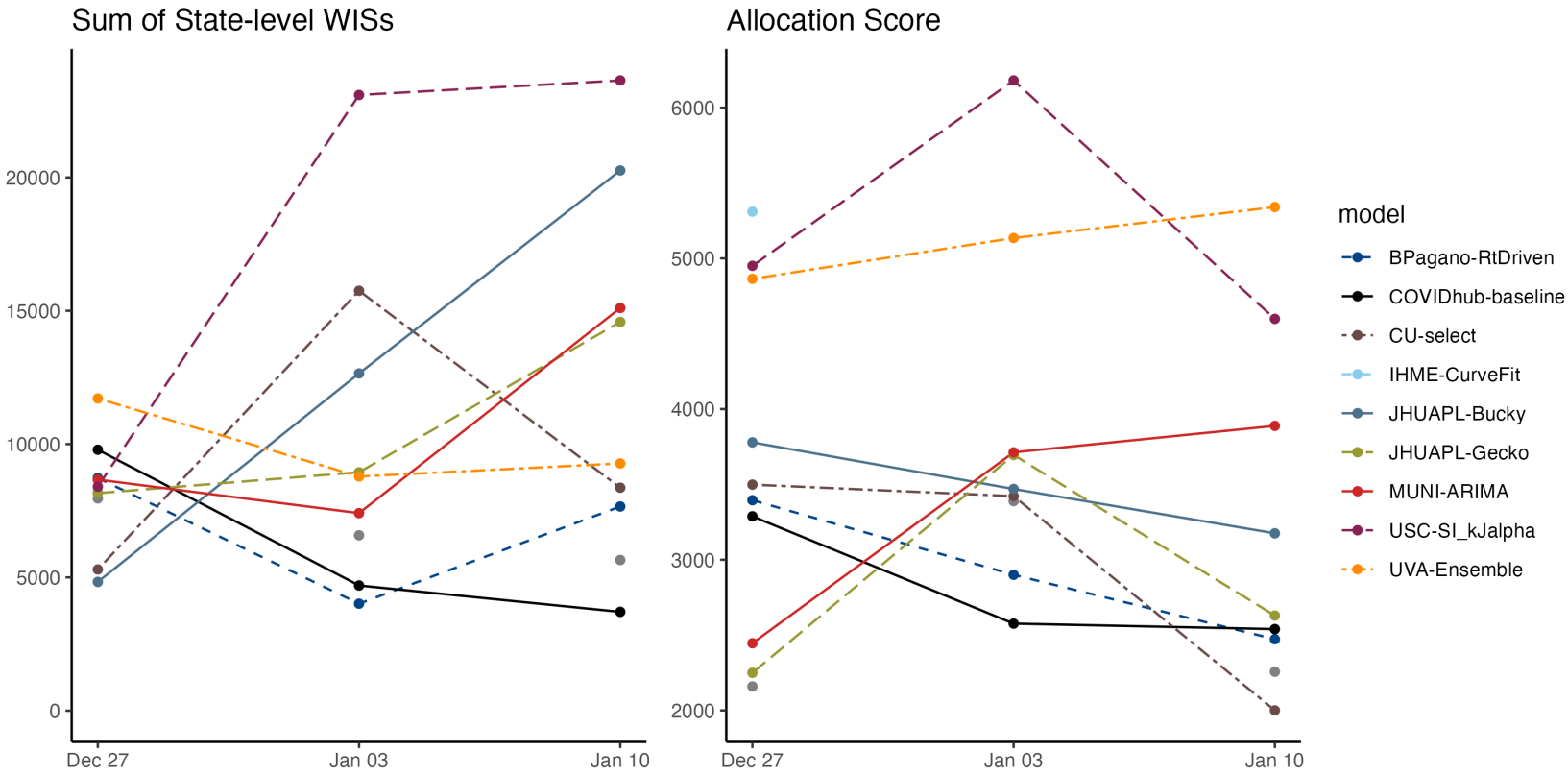


Some Observations

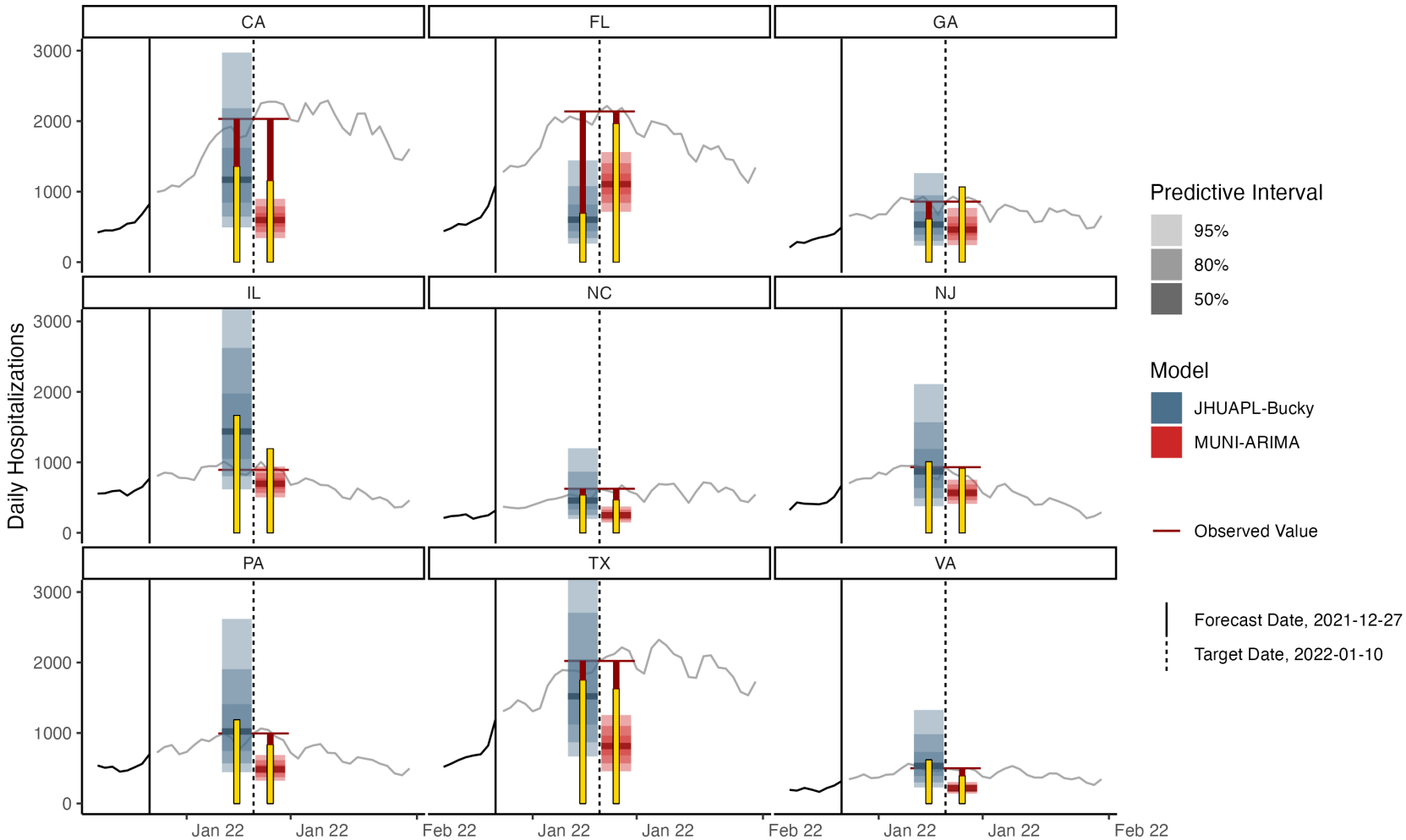
- related to the Murphy curves of Ehm, Gneiting, Jordan, Krüger, 2016
- extreme shortage or surplus diminishes oracle's advantage
- ranking consistent across large K region.



Alloscore and WIS rank models differently



Explanations?



Limitations

This is **post-hoc** analysis

Hub forecasters were unaware of

- an allocation score (on joint forecast)
- any allocation based loss
- our quantile interpolation/extrapolation methods (`distfromq`)
  - might be especially important for tails

We hope/think that allocation scoring is sensitive to implicit dependence structures in forecasts, but all work so far only refers directly to marginals - nothing yet with copulas, etc.

Thank you!

A very rough R package I wrote to implement scoring procedures:
https://github.com/aaronger/alloscore
A less rough package Evan wrote to implement cdf reconstruction
https://github.com/reichlab/distfromq