Infectious Disease Forecast Evaluation Based On Realized Social Utility:
Allocation Scores
MIDAS 2023

Aaron Gerding, Nick Reich, Ben Rogers, Evan Ray

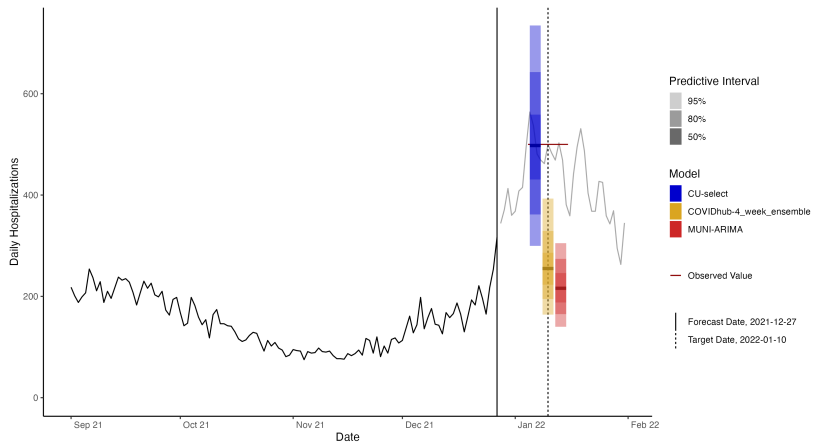School of Public Health and Health Sciences, UMass, Amherst

The Role of ID forecasting Hubs

Outbreak forecasting hubs developed to **"inform public health responses"**

Basic examples of **responses** are **allocation decisions** regarding

- ▶ medical supplies
- ▶ facility capacity
- ▶ personnel
- ▶ R&D funding
- ▶ public tolerance for NPIs

Basic Hub Products

Scoring rules and social welfare

A **scoring rule** S assigns a number $S(F, y)$ to the forecast F of Y when $Y = y$.

S if **proper** if $E_F[S(F, Y)] \leq E_F S(G, Y)]$ for all F, G

▶ ensures that honesty maximizes expected score

Hub goal of increasing social welfare by ranking and combining forecasts to best support PH responses seems to require use of proper scores.

Current standard is the **Weighted Interval Score** (a discrete version of the **Continuous Ranked Probability Score**)

▶ adopted for convenient scoring of quantile forecasts.

A central goal in design of scoring rules for social welfare

Create scenario where the forecaster believes maximizing her own expected utility aligns with leading the DM toward actions that maximize social utility.

Bayes Acts

Scoring rules can be cast in decision-theoretic framework of **Bayes acts**.

The **Bayes act** relative to a forecast $Y \sim F$ is the action x^F by which a DM will minimize **expected loss** $E_F[l(x, Y)]$.

▶ loss $l(x, y)$ is the **consequence** of experiencing y having prepared with x

Key point: $S(F, y) := l(x^F, y)$ is an *automatically proper* scoring rule.

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 $x_i^F = F_i^{-1}(\tau(K, F)), i = 1, \ldots, N$

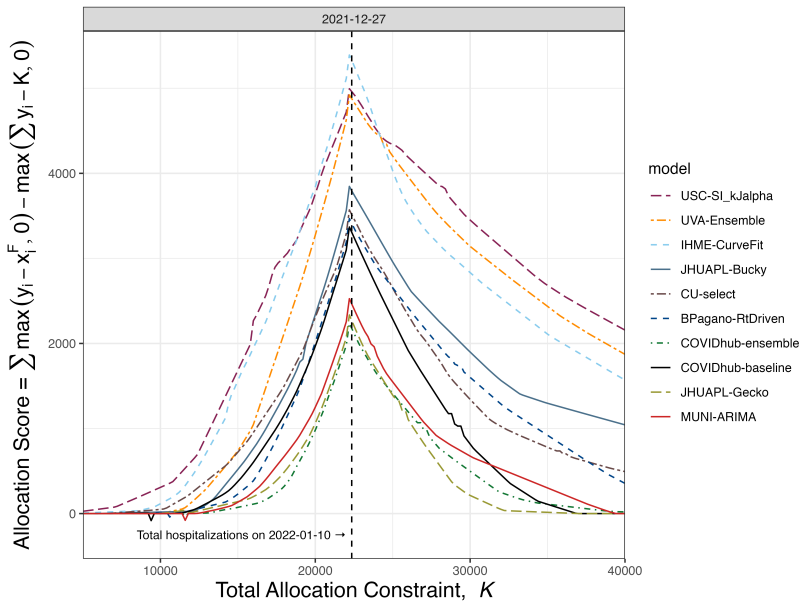▶ in practice, we find x_i^F's via an iterative method:
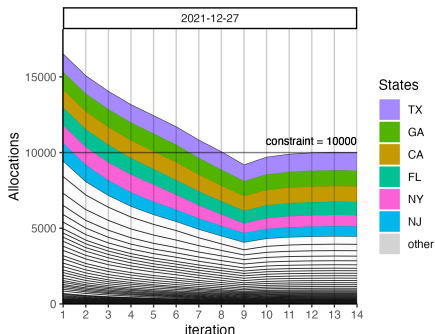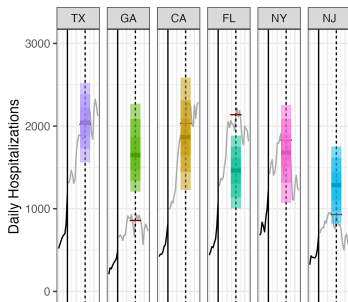
Application

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.
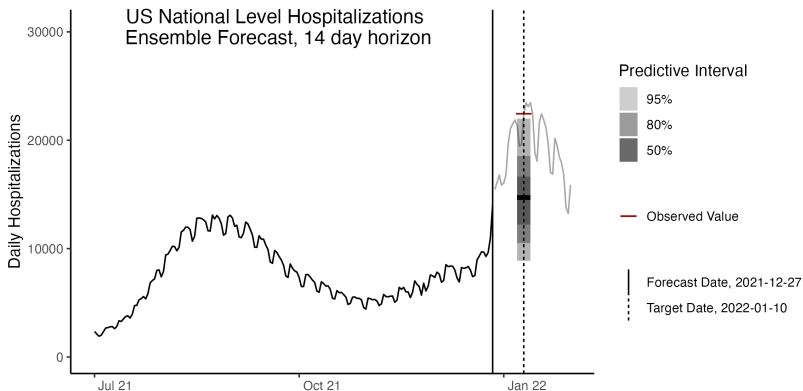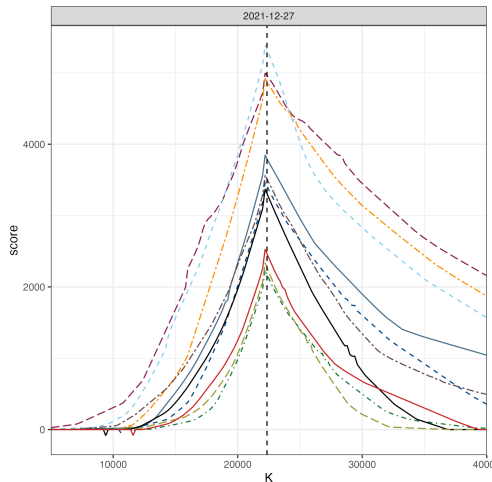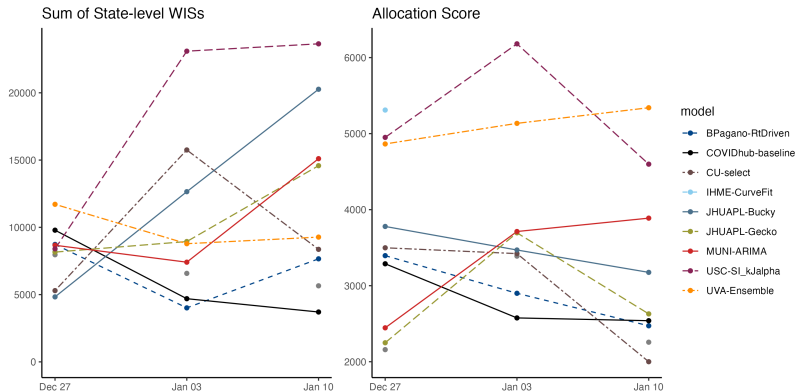
Oracle adjusted allocation scores near Omicron peak

Some Observations

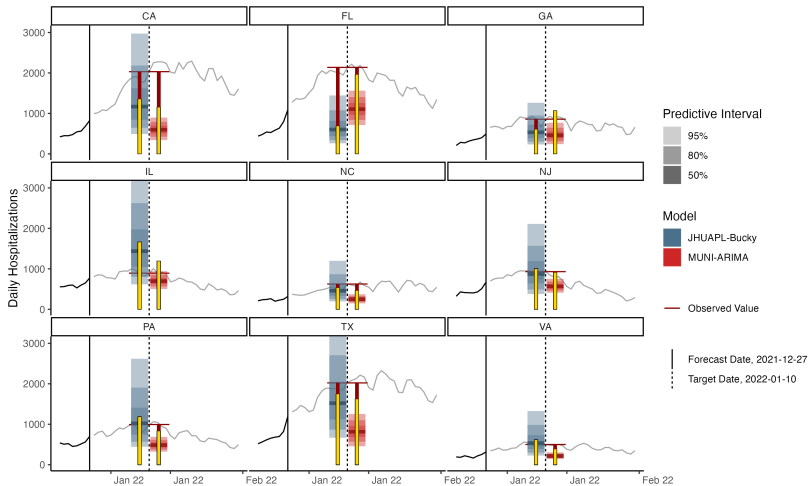- related to the Murphy curves of Ehm, Gneiting, Jordan, Krüger, 2016
- extreme shortage or surplus diminishes oracle's advantage
- ranking consistent across large K region.

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