Quantifying Epidemic Forecast Diversity and Change Through Optimal Transport

Lanyin Zhang
Dr. Srinivasan Venkatramanan

Background

- **Many modeling teams have created forecasts to predict cases, death, and hospitalization on COVID-19 Forecast Hub**
- **Multiple metrics are used to evaluate probabilistic forecasts but not intuitive for quantifying forecast diversity and change**
  - **e.g. COVID-19 Forecast Hub Ensemble uses Weighted Interval Score(WIS) to weigh forecast models**
  - **WIS measures how consistent a collection of prediction intervals is with an observed value instead of another distribution**
- **Optimal Transport and Wasserstein Distance (Earth Mover's Distance) provide a metric to compare two distributions and their geometry of the underlying space**
- **Compared to other metrics (e.g. KL distance and Jensen Shannon distance,) Wasserstein Distance is more intuitive in capturing qualitative changes**

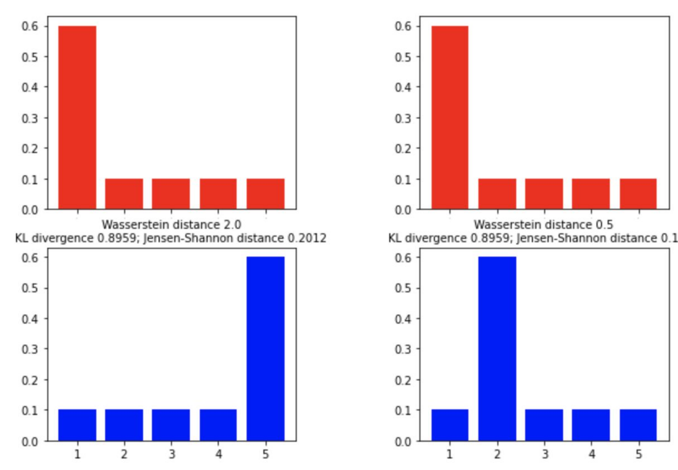


Figure #1 Distance Comparison

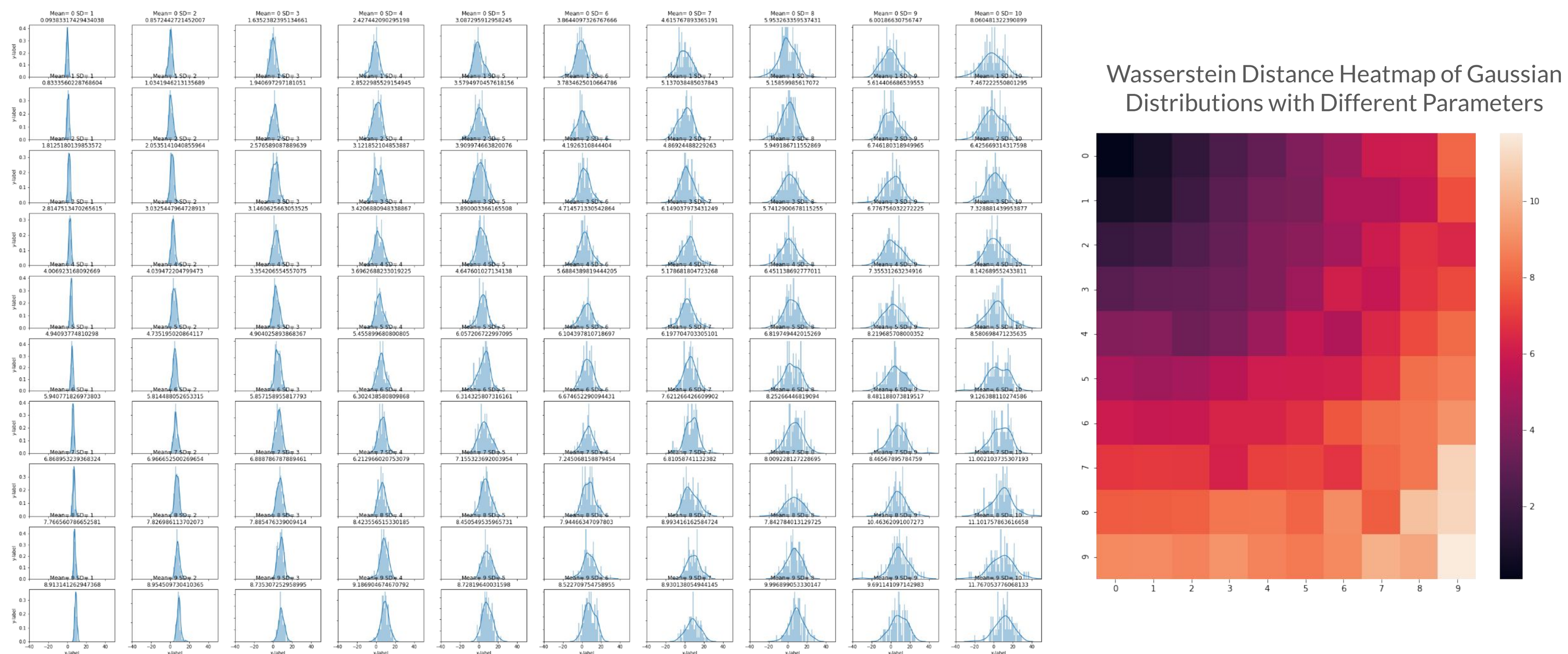


Figure #2 Gaussian Distributions Wasserstein Distance

Methods

- **Wasserstein Distance:**
  - **scipy.stats.wasserstein_distance calculates the first Wasserstein distance between two 1D distributions u, v:**

$$l_1(u, v) = \inf_{\pi \in \Gamma(u,v)} \int_{\mathbb{R}\times\mathbb{R}} |x - y| \mathrm{d}\pi(x, y)$$

- **Change: temporal updates of the same model**
  - **· Comparing forecasts for the same (target end date, location) across different horizons for the same model**
- **Diversity: similarity across different models**
  - **Comparing forecasts for the same (target end date, location, horizons) across different models**

Results - Change

- In general, longer time horizons tend to show larger uncertainty
- However, Omicron BA.1 wave created a large uncertainty when prediction time gets closer
- The absolute values of Wasserstein distance seem to be proportional to the ground truth
- Prediction before and after the peak have large Wasserstein distance

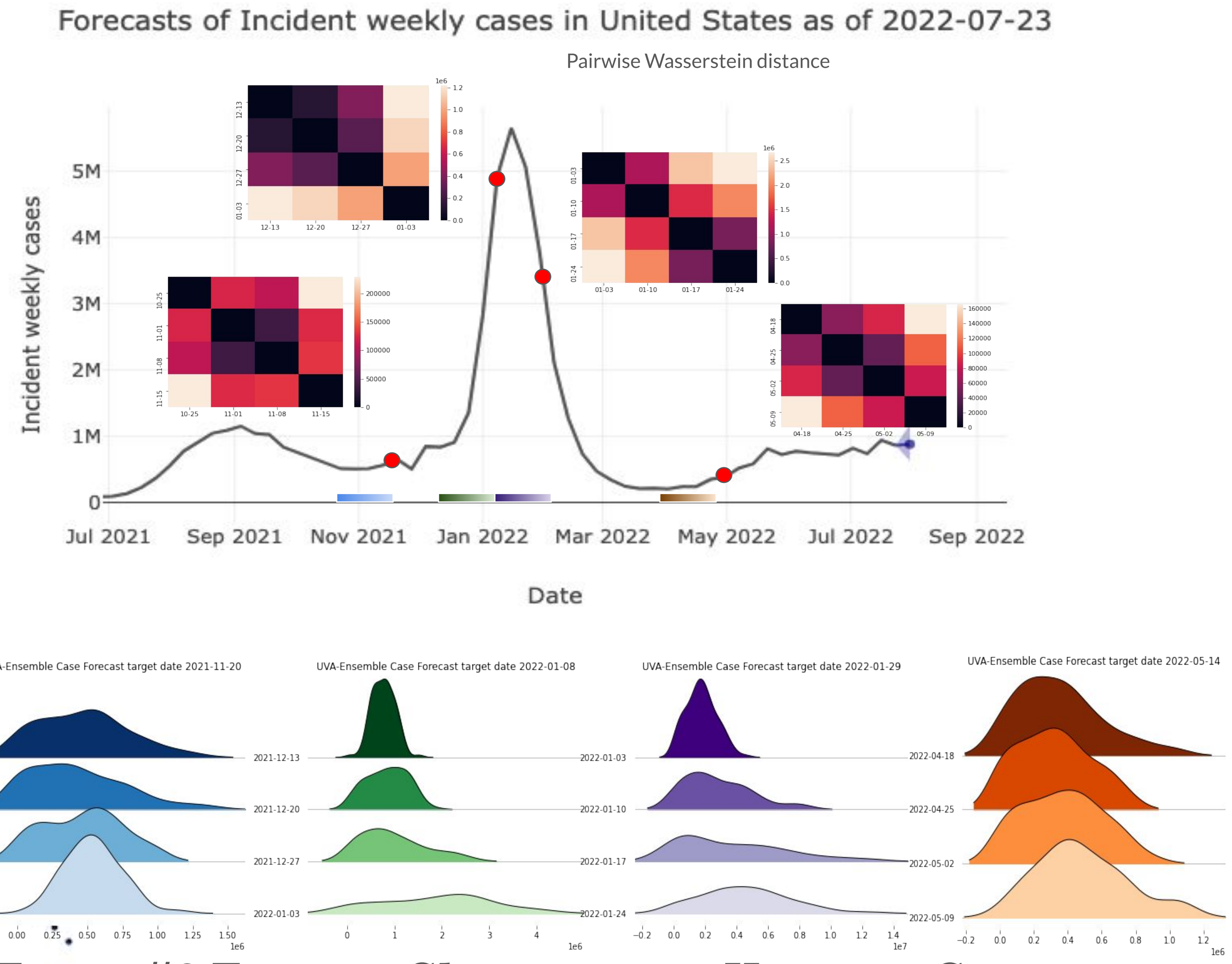
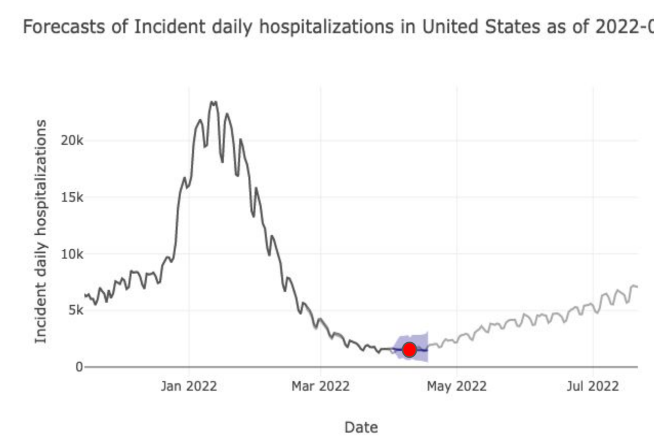
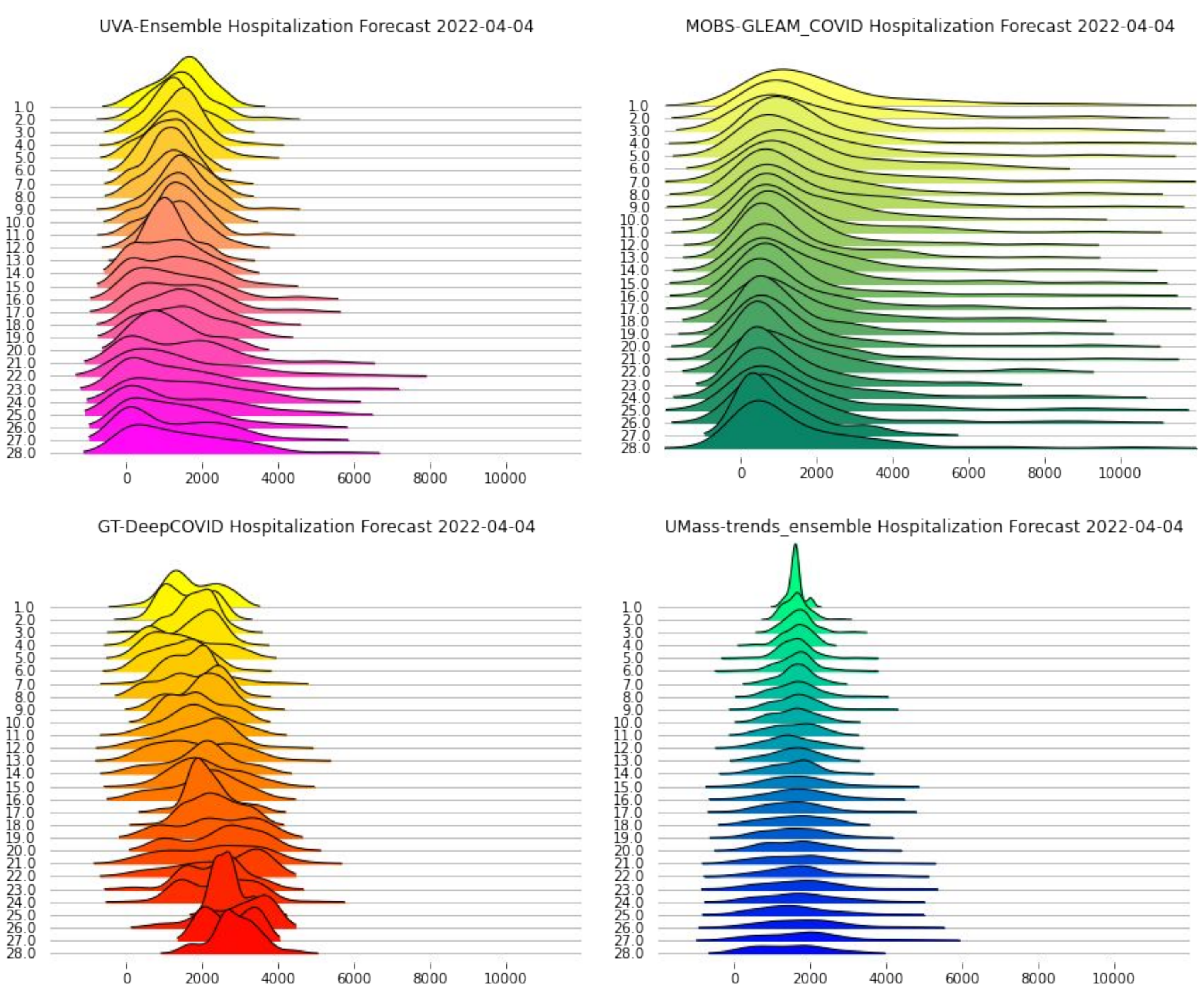


Figure #3 Forecast Change across Horizons Comparison

Results - Diversity

- Both mean and variance influence Wasserstein distance
- The most different model in day 1 is MOBS-GLEAM_COVID due to its prediction spread
- The most different model changed to GT–DeepCOVID due to its prediction of increasing hospitalization (others decrease in mean)
- UVA-Ensemble and UMass-Trend forecasts get closer





Heatmap of Wasserstein Distance for comparing different models

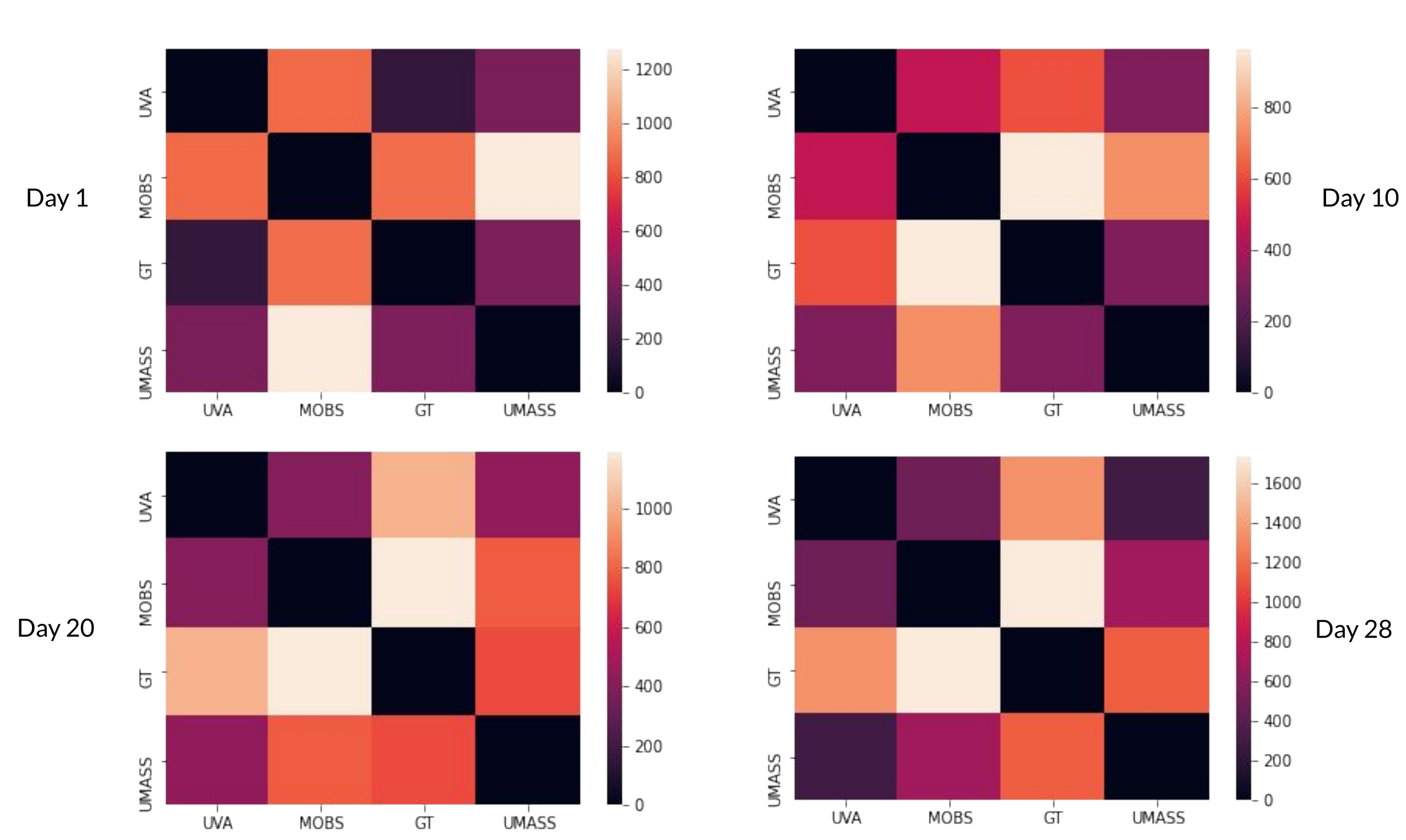


Figure #4 Forecast Diversity Comparison

Future Work

- **Normalize Wasserstein Distance across time and locations**
- **Use network-based methods to connect and compare distributions**
- **Model Ensemble with optimal transport methods, e.g. barycenter**
- **Evaluate geodesic and trajectory versions of Wasserstein distance**
- **Explore long-range projections (e.g. Scenario Modeling Hub)**

References

- **Cramer EY, Huang Y, Wang Y, et al. The United States COVID-19 Forecast Hub dataset. medRxiv. 2021. URL: https://www.medrxiv.org/content/10.1101/2021.11.04.21265886v1 References text in Franklin Gothic Book 18pt**
- **antike (https://stats.stackexchange.com/users/187743/antike), What is the advantages of Wasserstein metric compared to Kullback-Leibler divergence?, URL (version: 2018-06-13): https://stats.stackexchange.com/q/351153**
- **Cramer et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. URL: https://doi.org/10.1073/pnas.2113561119**

UNIVERSITY of VIRGINIA
BIOCOMPLEXITY INSTITUTE