COVID-19 EnsembleVis: Visual Analysis of County-level Ensemble Forecast Models

Sanjana Srabanti*  G. Elisabetta Marai†  Fabio Miranda‡
University of Illinois at Chicago  University of Illinois at Chicago  University of Illinois at Chicago

Figure 1: Enabling the visual analysis of COVID-19 ensemble forecast models with COVID-19 EnsembleVis. (a) We visualize the average error of the ensemble model at the county level. (b) Temporal distribution of the error of individual models, aggregated over all counties. (c) Distribution of errors over county population. Using the interface, we can notice that, in Week 42, the majority of the outlier counties are located in Texas (highlighted points in (c) and highlighted counties in (a)).

ABSTRACT
The spread of the SARS-CoV-2 virus and its contagious disease COVID-19 has impacted countries to an extent not seen since the 1918 flu pandemic. In the absence of an effective vaccine and as cases surge worldwide, governments were forced to adopt measures to inhibit the spread of the disease. To reduce its impact and to guide policy planning and resource allocation, researchers have been developing models to forecast the infectious disease. Ensemble models, by aggregating forecasts from multiple individual models, have been shown to be a useful forecasting method. However, these models can still provide less-than-adequate forecasts at higher spatial resolutions. In this paper, we built COVID-19 EnsembleVis, a web-based interactive visual interface that allows the assessment of the errors of ensembles and individual models by enabling users to effortlessly navigate through and compare the outputs of models considering their space and time dimensions. COVID-19 EnsembleVis enables a more detailed understanding of uncertainty and the range of forecasts generated by individual models.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Visual analytics;

* e-mail: ssrabta2@uic.edu
† e-mail: gmarai@uic.edu
‡ e-mail: fabiom@uic.edu

1 INTRODUCTION

The COVID-19 pandemic has upended the world over the past year, in a health crisis not seen since the 1918 flu pandemic. As of July 2021, the disease has caused the death of more than 4 million people worldwide, 600,000 in the US alone. While waiting for an effective vaccine amidst the surging cases worldwide, governments adopted measures to inhibit the spread of the disease. Despite these efforts, waves of COVID-19 infection with varying characteristics, continued to ravage communities [24], highlighting the necessity to understand the spatio-temporal complexity of the problem [25]. Forecast models are among the tools available to public health experts and policymakers to predict likely pandemic outcomes (i.e., number of cases, deaths) and prepare for different scenarios. In the past year, several models have been proposed by a myriad of experts, and using vastly different assumptions, methodologies, parameters and data sources. As a result, models produce a range of sometimes radically different forecasts, especially at higher spatial resolutions, limiting their use by decision makers and the public in general. In order to address this problem, more recently, ensemble models have been proposed for COVID-19 forecasts, with promising results [35, 44].

Ensemble models are usually created by combining different individual models (i.e., members), either through simple weighted averages or more sophisticated approaches, such as multiple linear regression and principal component regression [27]. Rather than relying on individual models, a set of forecasts in an ensemble indicates a larger range of possible future scenarios. Because of their high prediction accuracy, ensemble approaches are widely used in...
different domains, such as weather and climate [36], economy [47], and more recently in forecasting infectious diseases [35, 38]. Given the increasing importance of ensembles for COVID-19 forecasting, it is important to evaluate and compare ensemble members through space and time, as well as identifying potential limitations and guiding the improvement of both ensemble and individual models.

In this paper, we propose COVID-19 EnsembleVis, an open-source\(^1\) web-based interface to enable the visual analysis of the county-level differences between forecast and ground truth data of ensemble models and their members. We focus on fine geographic level data since it can reveal staggering disparities among different locations that were missed at coarser aggregation levels [22, 33], which signifies the need for even more precise predictions. By enabling effortless navigation through ensemble members and allowing the evaluation and comparison of multiple models, our proposed approach can be used to obtain a more detailed understanding of uncertainty and the range of forecasts generated by individual models, and highlight limitations of the models at the county level. This last point is a particularly important one, given the impact of the pandemic on different communities at different levels throughout the past year. Since COVID-19 ensemble models have only recently been shown to be effective [44], our work offers the first steps in eliciting some of the particular challenges related to the visualization and analysis of ensemble models of COVID-19 forecasts. We focus our efforts on the visualization of the outputs of ensembles and individual models. While our approach addresses some of these challenges, we believe that our work goes toward creating the necessary connections between ensemble visualization (a popular topic, especially in the weather domain) and pandemic ensemble models.

2 Related work

In this section, we briefly survey existing literature related to COVID-19 forecasting, a topic that has received significant attention recently due to the different ways the pandemic has impacted the world. Given how widely used ensemble models are in other domains, we also survey ensemble visualization techniques and systems.

COVID-19 forecasting. Over the past year, different research groups from academia and industry have been developing models to forecast COVID-19 cases and deaths. In the US alone, over 50 teams have made their forecasts publicly available, according to the COVID-19 Forecast Hub [5]. The majority of these models forecast both cases and deaths at a state level, while a smaller number forecast cases at the county level, (and one recent model also providing county-level fatalities [39]).

\(^1\)https://github.com/uic-evl/covid-19-ensemblevis

The vast majority of the models rely on the Johns Hopkins Coronavirus Resource Center confirmed case and death data [15], while a subset uses mobility [2, 4, 8, 13, 14, 18–20], weather [8, 13], US census [6, 18, 19], or hospitalization data [16, 20, 21]. Table 1 presents an overview of individual models. In April 2020, an initiative led by the Reich Lab at the University of Massachusetts started to collect and combine forecasts for US spatial units (states and counties), making the resulting ensemble data publicly available each week. This initiative was in close collaboration with the US Center for Disease Control and Prevention (CDC), who also release weekly ensemble forecasts as a real-time tool to help guide policy and planning [7].

Several interfaces and dashboards have also been made available with the goal of visualizing ensemble members [3, 9, 16, 18, 21], but they are restricted to individual members or lower spatial resolutions. COVID-19 EnsembleVis is the first visual interface that allows users to compare and analyze ensembles and individual members at the county level.

**Ensemble visualization.** Ensemble data is common in several domains, such as biomedical images [34], network security [32], climate simulations [40], machine learning [46], and due to the complexity of the data, ensemble visualization faces a variety of research challenges, such as scalability, extraction of trends, differences and commonalities [42]. Previous work has focused on proposing frameworks to support visual analysis of ensemble data through a combination of statistical visualization techniques and user interaction [43], visual glyphs, such as radar plots [37], ribbons and spaghetti plots [45], clustering [30, 31] and probabilistic [28, 29] and trend analysis [41]. A complete survey of visualization and visual analysis of ensemble data can be found at Wang et al. [48]; the survey focuses on more common ensemble data, such as weather and climate simulations, and lists a series of six common ensemble visualization tasks: overview, comparison, clustering, temporal trend analysis, feature extraction, and parameter analysis. In our work, we use a subset of these tasks to visualize COVID-19 forecast ensemble models, with the goal of indicating and starting to lay down bridges between ensemble visualization research and pandemic predictions.

### Table 1: Overview of the individual models considered by our tool.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Team</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU-TimeSeries [10]</td>
<td>CMU</td>
<td>JHU case and death counts</td>
</tr>
<tr>
<td>CU-Select [20]</td>
<td>Columbia U.</td>
<td>JHU case and death counts, hospitalizations, ICU admissions, mobility</td>
</tr>
<tr>
<td>FAIR-NKAR [8]</td>
<td>Facebook</td>
<td>NY Times case counts, weather, mobility</td>
</tr>
<tr>
<td>IBM_MED_CovidProject [9]</td>
<td>JHU</td>
<td>JHU case and death counts</td>
</tr>
<tr>
<td>IowaStateL.W-STEM [3]</td>
<td>Iowa State</td>
<td>NY Times, Health department, county-level infected and death cases</td>
</tr>
<tr>
<td>JHU/APL-Backy [14]</td>
<td>JHU</td>
<td>JHU, hospitalizations, mobility</td>
</tr>
<tr>
<td>JHU_CDCovidSP [19]</td>
<td>JHU</td>
<td>JHU, US Census (population, mobility)</td>
</tr>
<tr>
<td>LANL-GrowthRate [16]</td>
<td>Los Alamos</td>
<td>HHU, population</td>
</tr>
<tr>
<td>LNQ-esn1 [17]</td>
<td>SAS</td>
<td>JHU</td>
</tr>
<tr>
<td>OneQuestNight-Ml [4]</td>
<td>JHU</td>
<td>mobility</td>
</tr>
<tr>
<td>PandemicCentral-USCounty [18]</td>
<td>Pandemic Central</td>
<td>JHU case and death cases, US Census, CCVI, mobility</td>
</tr>
<tr>
<td>UCLA-SnEIR [21]</td>
<td>UCLA</td>
<td>JHU case and death cases, hospitalizations</td>
</tr>
</tbody>
</table>

3 COVID-19 Forecasts and Ensembles

An initiative led by the University of Massachusetts has created a central repository, called COVID-19 Forecast Hub, that collects and organizes submissions with COVID-19 forecasts in the US, both at the state level and at the county level. These submissions are developed independently and shared publicly. Submissions can include week-ahead forecasts of COVID-19 deaths and/or cases following the CDC’s epidemiological weeks. For example, during epidemiological week 1 (EW1), a team can submit a forecast for the
The CDC recently highlighted how important it is to bring forecasts and geographical unit, the first file lists the models’ weights as well as ensemble forecasts, are made available as CSV files with a weekly temporal resolution and state and county spatial resolutions. Ground truth data is also collected by the Forecast Hub, allowing the straightforward computation of error metrics. To be more precise, the error of a model can be calculated as the difference \( e_{ct} = y_{ct}^c - x_{ct}^c \), where \( y_{ct}^c \) is the model’s predicted value for county \( c \) at week \( t \), and \( x_{ct}^c \) is the ground truth for county \( c \) at week \( t \). This paper, positive differences (i.e., predicted value greater than ground truth) are represented by shades of blue, and negative differences by shades of red.

On top of the individual forecast CSV files, the central repository also hosts a weight CSV file and an eligibility CSV file. For each week and geographical unit, the first file lists the models’ weights used in the computation of the ensemble forecast, and the second file informs whether or not a given model is eligible for inclusion in the ensemble forecast. Both files are updated weekly.

It is important to note that UMass’ center interprets forecasts as unconditional predictions about the future. In other words, predictions should consider uncertainty across a wide range of future scenarios (e.g., new social distancing mandates), and it is up to the team to select and submit one or a combination of predictions across the most likely scenarios. In this work, we make use of UMass’ COVID-19 Forecast Hub ensemble model, which combines the individual models highlighted in Table 1. We focus our attention on the visualization of 2-week-ahead predictions.

4 Requirements

The CDC recently highlighted how important it is to bring forecasts together to help understand how they compare with each other and how much uncertainty there is about what may happen in the future [7]. Considering that, we identified three main tasks that can be facilitated by a visual interface: 1) provide an overview summary of the errors of the models to assess uncertainty; 2) identify spatiotemporal trends, i.e., how a group of members changes over space and time; and 3) visually identify differences between two or more ensemble members. In order to accomplish these tasks, we identified the following requirements for our visual interface:

[R1] Support the identification of spatiotemporal patterns. Explore the spatial and temporal patterns of one or more ensemble members to identify regions or periods with above average forecast error and uncertainty.

[R2] Support the comparison of ensemble members. Compare predictions of different ensemble members to evaluate the spatiotemporal performance of different models.

[R3] Support the analysis of the relationship between model performance and sociodemographic features. Explore the relationship between sociodemographic variables and model performance to identify regions not adequately represented in the model.

5 COVID-19 EnsembleVis

In order to satisfy the previously detailed requirements, we built COVID-19 EnsembleVis, a web-based interface to facilitate the visual exploration of COVID-19 forecast ensembles. The interface is composed of three main views: spatial view, temporal view, and distribution view. These three views are all linked, such that a selection in one view will highlight the appropriate data in other views. In this work, we consider weekly predictions between April 6, 2020 and July 5, 2021.

Spatial view. This component is composed of a map with US counties (Figure 2) and enables the identification of spatial patterns (R1). Each county is painted according to the average ensemble member’s forecast error over all weeks (i.e., \( \sum_{\text{weeks}} e_{ct}^c \), for each county \( c \)). A divergent color scale is used to indicate negative and positive forecast errors. In order to enable detailed analysis, the user can select an individual county, and the temporal view will update accordingly.

Temporal view. This view uses a matrix heatmap to display the weekly average error for each ensemble member (Figure 3). Each cell is painted according to the average ensemble member’s forecast error over all counties (i.e., \( \sum_{\text{counties}} e_{ct}^c \), for each week \( t \)). The temporal view allows the visualization of forecast errors over time (R1) and the comparison of individual ensemble members (R2) that can help in the identification of weeks with bad predictions (e.g., predictions with high errors) from ensemble members. The temporal view is linked with the spatial view. If a single county is selected, the weekly average error for each model of that particular county will be shown (Figure 6). Furthermore, the temporal view enables the user to add temporal constraints by selecting specific weeks of interest, which will update the other views. The same divergent color scale from the previous component is also used in the temporal view. Missing predictions are displayed in light gray.

Distribution view. This view incorporates a 2D scatterplot that shows, for each week, the distribution of per-county ensemble prediction errors (normalized by population) by the log of the total county population. Each data point then represents the ensemble prediction error for a given county at a given week. Users can use the scroll wheel to view different weeks or select specific weeks from the temporal view. In this version of the interface, we only consider the population information, partially meeting R3. The main
The purpose of the scatterplot is to assess any possible relationship between prediction error and demographic variables. The distribution view also allows the user to brush the scatterplot and select data points of interest; this will then update the other views, enabling the assessment of the spatiotemporal distribution of the errors.

**Implementation.** The COVID-19 EnsembleVis interface was implemented using Angular, D3, and TopoJSON so that it could be easily accessed through a web browser. We collected the predictions from UMass’ COVID-19 Forecast Hub and pre-processed them using Python and Jupyter Notebook. The EpiWeeks Python library was used to compute epidemiological weeks (following CDC’s standard). The pre-processing step is responsible for aggregating the data over counties and epidemiological weeks. The pre-processed data is then stored as a single JSON file and accessed by the front end. The source code and pre-processing stages are available at https://github.com/uic-evl/covid-19-ensemblevis.

**6 Exploring Ensemble Predictions**

In this section, we illustrate an initial case on how our tool can be used in the visual analysis of ensemble models, with a focus on understanding how prediction errors are spatially distributed over different counties. We begin the exploration by visualizing the distribution of ensemble prediction errors by county population using the distribution view (Figure 1(right)). We notice a particular week (Week 42) where certain counties present a larger prediction error. We select these data points in the scatterplot and notice that most of them are counties in Texas. While it is not possible to precisely say why this happened, such visualization can foment discussions on the shortcomings of current modeling practices, such as data availability, parametrization for those counties, etc.

We also used COVID-19 EnsembleVis to explore the spatial distribution of errors in Week 47, which showed another unusual pattern (Figure 5). After selecting a set of points in the scatterplot (left side of the figure), we noticed that these points are from counties in Missouri (right side). Unlike the previous example, however, we were able to pinpoint the source of such an unusual pattern: on March 8, 2021, the Missouri Department of Health and Senior Services updated the number of daily cases to show 81,206 previously unreported infections.

**7 Conclusion and Future Work**

COVID-19 EnsembleVis is a visual interface built specifically for the analysis of COVID-19 forecast ensemble models. By using three different visualization components, we enable the investigation of both the ensemble and individual models, from both spatial and temporal perspectives. COVID-19 EnsembleVis takes the first steps toward the visualization of county-level forecast ensemble models. Although only presenting a collage of known visualization metaphors, this work creates a foundation that will enable researchers to apply ensemble visualization techniques to efforts related to the recent pandemic. We hope this can help foment a discussion that builds bridges between ensemble visualization researchers (mostly focused on weather and climate data) and modeling and public health experts. Additionally, we believe there are several interesting challenges that could certainly benefit from collaboration between these fields. For instance, while we focus our efforts on the visualization of the output of the models, certain research teams do make their code publicly available, opening doors to have a more vivid picture of their computational mechanisms. Therefore, understanding the impact of different parametrizations through visual analytics tools is a path that can not only increase performance and accuracy but also increase public trust in these models.

We also believe that there is an opportunity to use metaphors that were previously introduced by visual analytics tools to visualize weather and climate ensemble data. Given the potential impact of COVID-19 predictions, in future work we will also investigate how predictions can better inform policymakers and their decisions from a visual analytics perspective [23]. Furthermore, we will explore the relationship between models and data sources. As shown in Table 1, models use different data sets, and this can heavily impact the predictive power of both individual members and ensembles. Given that, we believe it would be interesting to understand the relationship between data sets, model performance, and social demographic variables. Moreover, given the number of possible data slices, it would be important to guide the user in the exploratory process and highlight potentially interesting data features. We will also extend COVID-19 EnsembleVis to enable the visualization and exploration of COVID-19 forecasts from other regions of the world, including Europe [12].
REFERENCES