2023 IEEE VIS Workshop on Visualization for Pandemic and Emergency Responses (Vis4PandEmRes) | 979-8-3503-3026-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/Vis4PandEmRes60343.2023.00006

Visual Analytics based Search-Analyze-Forecast Framework for Epidemiological Time-series Data

Tuna Gonen* University of Oxford
Yiwen Xing† King's College London
Cagatay Turkay‡ University of Warwick
Alfie Abdul-Rahman§ King's College London
Radu Jianu¶ City, University of London
Hui Fang‖ Loughborough University
Euan Freeman** University of Glasgow
Franck P. Vidal†† Bangor University
Min Chen‡‡ University of Oxford

Abstract

The COVID-19 pandemic has been a period where time-series of disease statistics, such as the number of cases or vaccinations, have been intensively used by public health professionals to estimate how their region compares to others and estimate what future could look like at home. Conventional visualizations are often limited in terms of advanced comparative features and in supporting forecasting systematically. This paper presents a visual analytics approach to support data-driven prediction based on a search-analyze-predict process comprising a multi-metric, multi-criteria time-series search method and a data-driven prediction technique. These are supported by a visualization framework for the comprehensive comparison of multiple time-series. We inform the design of our approach by getting iterative feedback from public health experts globally, and evaluate it both quantitatively and qualitatively.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

The ongoing COVID-19 pandemic has seen a surge in the use of time-series data depicting both national and international core disease statistics, such as the number of cases, hospitalizations or deaths, as well as data on measures and interventions such as the number of tests and vaccinations administrated. Dashboards with a national (e.g., The UK government's coronavirus dashboard [11]) or an international focus (e.g., [1, 22]) have stood out as primary interfaces where experts and the public alike use to stay informed on the progress of the pandemic. As observed in our ongoing collaborations with epidemiologists, such dashboards have been used by experts to estimate how their country/region compares to others and estimate what the near future could look like at home. And as evidenced in our own empirical research (detailed later in the paper), this kind of "predictive" use of data from "similar" locations is even more pronounced in countries that have limited capacity in rapid epidemiological modeling. There is, however, a lack of visualization approaches that can support such data-intensive search and forecasting processes.

*e-mail: tuna.gonen@spc.ox.ac.uk
†e-mail: yiwen.xing@kcl.ac.uk
‡e-mail: cagatay.turkay@warwick.ac.uk
§e-mail: alfie.abdulrahman@kcl.ac.uk
¶e-mail: radu.jianu@city.ac.uk
‖e-mail: h.fang@lboro.ac.uk
**e-mail: euan.freeman@glasgow.ac.uk
††e-mail: f.vidal@bangor.ac.uk
‡‡e-mail: min.chen@oerc.ox.ac.uk

This paper aims to address this gap and proposes a visual analytics approach where short-term forecasts of pandemic progress at a location can be made based on a carefully curated selection of similar locations. More specifically, at its core, this paper is asking "how can short-term predictions (on the progress of the pandemic) of a particular location could be made based on other locations that showed similar disease progression patterns at some point in the past?". There are, however, a number of challenges towards this aim: it is challenging to identify effective ways to define and compute similarities between the time-series for the purposes of search, it is unclear how suitable indicators and locations for forecasting could be identified, and there is limited work in generating and presenting data-driven predictions based on similarity search results.

A robust similarity quantification is the core pillar of an effective similarity search algorithm, however, there is no single established notion of similarity for time-series. Several metrics to quantify similarity within time series [7] exist and it is rarely clear which metrics are the most suitable for a given context. Moreover, choosing a time-window that appropriately supports analytic goals and returns relevant similarities complicates matters further. Such time-windows could span from a few weeks to the entire duration of the pandemic and the degree to which two regions exhibited similar epidemiological patterns would vary greatly depending on this choice.

An effective data-driven prediction such as the one we envision requires a carefully selected list of countries that are representative of a target location and can be used as a basis to predict its epidemiological evolution in the near future. However, estimating the progress of a disease is a complex task as multiple factors, such as levels of vaccination, seasonal factors, population demographics, or a population's response to public health interventions, are concurrently at play. Since such complexities are difficult to model and integrate into automated processes, the search and refinement of regions that could inform predictions often benefit greatly from the involvement of human experts and their tacit knowledge.

As a response to these challenges, we present a visual analytics approach to support human-guided and data-driven predictions through a search-analyze-predict process. It integrates a multi-metric, multi-criteria interactive search process and a data-driven prediction technique into a visualization framework designed to enable a comprehensive comparison of many time-series. We informed the design of our approach through iterative feedback from public health experts globally and implemented our ideas within a visualization tool called VASAP. This was deployed as part of the RAMPVIS infrastructure [5] that aims to collate general COVID-19 related visualizations with specialized epidemiological and public health tools. We evaluated VASAP both quantitatively and qualitatively via user studies and interviews with domain experts.

2 Related Work

VA for time-series analysis & forecasting. Mining of time-series data involves various analytical tasks such as classification, anomaly detection, clustering, search and prediction in fields from finance to human behavior. Many studies have been done on mining of

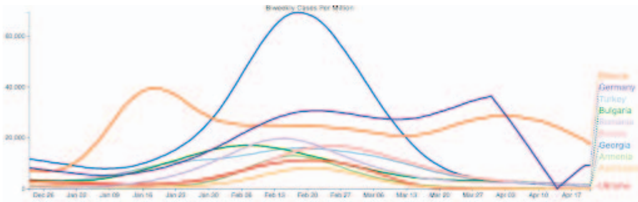

Figure 1: Commonly-used time series plots, such as the one shown here, cannot provide healthcare experts with adequate support in their attempts to predict the unknown trend of a target location based on the known trend in other locations. It would be difficult for the healthcare experts to determine if the reference time series were appropriately selected, identify and align matched periods in different time series, or mentally estimate the trend of the target location.

time-series data [8]. For longer time-series, with increased data-size and computational expense, dimensionality reduction techniques are used to increase the efficiency of algorithms [13].

The analysis of time-series data has fueled a consistent and long-lasting research drive for ways to visualize and make sense of such data with a wide range of static and dynamic visualization techniques [3]. Due to our interest in showing trends and comparing time-series, our focus is on designs that draw on multiple line charts [23] and lens-based interaction for time-series visualizations [16].

VA for epidemiology. Visual analytics systems have already supported data analysis within the public health domain, in particular in the context of epidemiological data [20], and systems level indicators [6]. Similarly, visualization approaches have been employed both for pandemic management, understanding and control, as well as decision support tools to inform experts [24], while also studying visualisations from both perception of time-series similarity [9] and prediction [18] perspectives.

In the context of COVID-19, visual analytics systems supporting situational awareness and decision making have been proposed. LF-peers [4] allows users to detect similar epidemiological movements occurring in the same period and to explore posterior variations using other variables. CoronaVis [12] focuses on visualizing hospital capacities on a map with other variables with glyphs. Our work builds on and advances these contributions by introducing time-period search and prediction capabilities.

3 Background and Approach

Since December 2019, the world has been facing an unprecedented health and economic crisis. Many countries have built their policy management approaches based on data surveillance (e.g., [19]). While many developed countries have relied extensively on epidemiological modeling (e.g., [17]), the number of developing countries that are able to access epidemiological modeling capabilities has so far been limited [2].

RAMP VIS started in May 2020 as a group of 22 volunteers specialized in data visualization and visualization, who answered a call to support the modeling scientists and epidemiologists in the Scottish COVID-19 Response Consortium (SCRC) [1]. While the main mission of the RAMP VIS activities has been to support the development of epidemiological models, the group anticipated the needs for various visual analytics techniques for processing time series in analytical and model-developmental visualization [5].

We recognized that healthcare experts in many low- and middle-income countries (LMICs) were acquiring latest information about COVID-19 from the public data sources and did have expert knowledge about healthcare matters (e.g., behaviors of COVID-19 variants, vaccination status) and regional factors (e.g., geographical, demographic, social and infrastructural), but struggle to access modeling capacity in the short-term, as also recognized by the epidemiological community [2]. If healthcare experts in LMICs could easily find historical COVID-19 data in other regions that exhibit similar patterns to the current data in a *target* region, they would be able to use their expert knowledge to filter out less relevant past data. Based then on the most relevant past data, healthcare experts could anticipate what may happen next in the target region concerned and make better preparation. Built on this idea and the initial software prototype, we developed our deplorable tool, VASAP, and made it a part of the RAMPVIS infrastructure (`https://sites.google.com/view/rampvis`).

[1] https://scottishcovidresponse.github.io/

3.1 Tasks and Requirements Analysis

Anticipatory task and requirement analysis: We anticipate the basic workflow for making data-driven predictions as follows: (i) given a time series of a *target* region, a user needs to select a set of *reference* time series from the data of other regions. (ii) once a set of reference time series are identified, the user then defines a time window of the target time series and attempts to identify periods in the reference series that match the time window. (iii) even if we assume that the user is satisfied with the matching periods determined visually, the user would still need to determine a section of time series succeeding the matching period for each reference time series. Such a succeeding section can be used as a piece of *baseline* data to help the user about the would-be succeeding section of the target region. Since there are a number of reference time series, the user needs to align all such succeeding sections to create an ensemble of baseline data to aid the prediction.

Initial prototype evaluation and survey of needs: We developed an initial web-based prototype based on the above anticipated tasks which we then demonstrated to two domain experts in a LMIC and made adjustments (as discussed in the Evaluation section). Once the initial prototype is ready, we involved further domain experts in LMICs in the development lifecycle (e.g., tasks and requirements analysis, testing, and evaluation) by conducting an engagement survey with 44 experts in LMICs. These activities will be further detailed in the Evaluation section.

Consolidating Analytical Tasks: Following the initial expert consultation and engagement survey, we outlined the following tasks to be supported by a visual analytics tool:

T1: *Searching for an initial list of reference time series.* Given a set of user-defined criteria including a time window in the target time series, this task is to use an algorithm to search in a large collection of time series and identify those featuring one or more periods that match with the target time window.

T2: *Identifying the list of reference time series.* Given an initial list of reference time series, this task is to remove those time series that are considered to be inappropriate for supporting the prediction task to follow. Visualization is used to support this task by (i) making the observation of individual time series easier, (ii) allowing the user to use knowledge beyond the data being visualized, and (iii) enabling the visualization of additional data to strengthen the filtering decision based on (i) and (ii).

T3: *Forecasting the near future based on the selected baseline reference time series.* Given a filtered list of reference time series, this task is to allow the user to visualize sections of baseline data for prediction, and a mean curve as the suggested prediction.

3.2 Similarity Measures for Comparing Time Series

There are many ways to quantitatively compare time-series [7]. Broadly, we can consider two types of comparison metrics: (i) dissimilarity metrics, often based on the notion of "distance", and (ii) similarity metrics, e.g. measuring a correlation. There is no single best metric as their relative performance often depends on the properties of the data. It is one of the key reasons why an interactive tool must offer at least a limited selection of comparison methods.

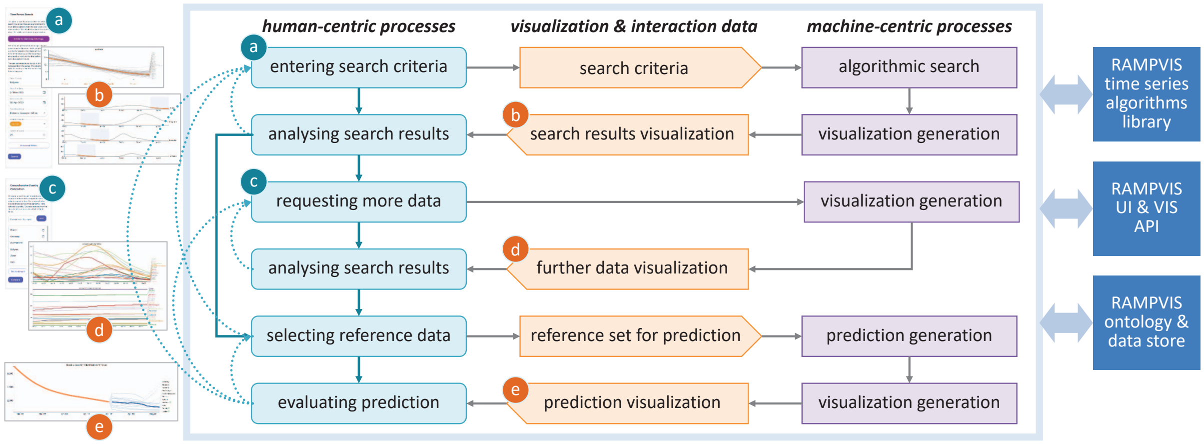

Figure 2: A typical workflow of the search-analyze-forecast tasks and their relation to the components of the RAMPVIS infrastructure, including the time series algorithm library, UI and VIS API, ontology, and data store. The letters match the process step and supporting visualizations.

The Manhattan distance, also known as L_1-norm, has been traditionally popular in computer science due to the use of absolute values instead of squared values. This is particularly true when time of execution is an issue. It also makes the Manhattan distance relatively robust against outliers compared to the Euclidean distance (also known as L_2-norm). Indeed, the use of squared values gives a higher weight to large discrepancies. Chebyshev, also known as L_∞-norm, returns the largest pair-wise discrepancy between the two vectors. It makes it unintuitive compared to the Manhattan and Euclidean distances, and unpractical if the data is noisy (unless the data has been smoothed as a pre-processing step). None of the three distances discussed above are scale invariant. To overcome this, it is important to "normalise" the data as a pre-processing step. Other metrics such as Pearson's correlation, aim to address this deficiency. Bray–Curtis dissimilarity (often erroneously called a distance) is a weighted version of the L_1 norm. It is, however, bounded between 0 and 1. If the data in two vectors $\mathbf{x}$ and $\mathbf{y}$ are positive (as it is the case with the time-series considered here), the Bray–Curtis dissimilarity is equal to the Canberra distance.

In addition to the distance-based metrics, two algorithmic elastic-based measurements, including Dynamic Time Warping (DTW) and Longest Common Subsequence (LCS), are set as recommended measurements in the system. DTW aligns time-series data before calculating the distance, which allows similar patterns to be retrieved even when lags are present. Unlike DTW which maps all points between pairs, LCS searches similar sub-sequences of the pairs so that it is more resistant to outliers. These two algorithmic measurements provide metric alternatives to deal with common time-series matching issues such as has lagged effects or misalignment. Notably, similarity based metrics are not selected as recommended distance measurement methods in our system since these metrics favor normalized data. We are more interested in retrieving data with similar magnitude over highly correlated time-series.

4 Search, Analysis, and Prediction

VASAP is a visual analytics tool for enabling the search, analysis, and prediction workflows as illustrated Fig. 2, and it is designed to meet the requirements discussed above. It is part of the RAMPVIS infrastructure [14], which consists of a large collection of dashboards and plots [15] and several interactive tools including VASAP.

The VASAP tool consists of four main software components. The frond-end component support users with interactive visualization as shown on the left of Fig. 2. The back-end components are *a time series algorithm library*, *an API for user interface and visualization plots*, and *data access ontology and data store* as shown on the right of Fig. 2. The Our World in Data COVID dataset[2] is chosen as the data source for obtaining a comprehensive collection of epidemiological time-series data that is updated daily. The data is automatically downloaded onto the RAMPVIS infrastructure every day according to a scheduler agent, which also organize the data into time series data streams that can be retrieved according to locations (e.g., countries or regions) and types (e.g., cases, deaths, etc.). To ensure consistency of data between different regions, time-series with missing data are excluded when daily data is used and biweekly and weekly smoothed curves are included in the system.

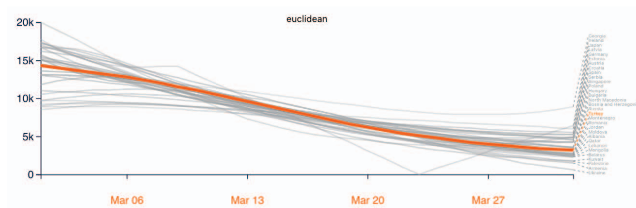

Figure 3: A visualization displays the top $N = 30$ patterns that are most similar to the target pattern shown in orange. The time axis corresponds to the target pattern. All other patterns are aligned to the target pattern in order for users to assess the quality of the search, before analyzing individual search results.

VASAP is equipped with a web-based user interface designed to reflect the three tasks described earlier. It enables users to perform their search, analysis, and prediction tasks with the aid of algorithms and visualization, while allowing the users – through interaction – to bring their knowledge into the workflow by inputting their preferences and constraints into the search and analysis processes.

4.1 Time-Series Search

With the VASAP tool, the users can specify a time window for the target country or region and the type of data streams.

VASAP uses the selected similarity measure to search among all time series in the search space for data patterns similar to the time window of the target region. There are many pattern matching methods in the literature. We studied and implemented and tested 12 time series comparison methods and selected 6 of them to be included in VASAP as discussed in 3.2. Considering that most users may not have in-depth knowledge about similarity measures, VASAP provides users with a default measure (Euclidean) as well as a help page explaining all the available measures.

[2] https://github.com/owid/covid-19-data/

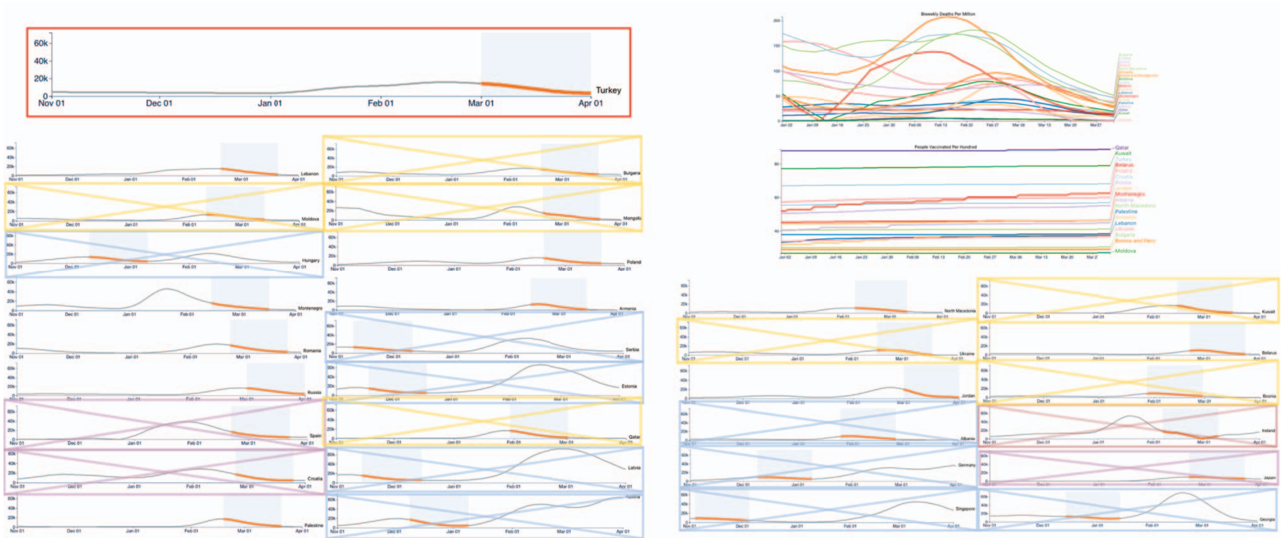

Figure 4: The time series of the target region (the top left plot), and the set of top N search results as the initial set of reference time series. The matching pattern in each time series is highlighted with the $N+1$ plots organized vertically in the order of similarity to help users assess the suitability of individual reference time series for prediction.

The VASAP user interface allows users to refine the search space by using advanced filters such as specific continents, a specific population size, and specific time period.

As soon as the search is complete, VASAP displays the top N patterns that are most similar to the time window of the target region. Fig. 3 shows top 30 patterns that match the target time window of a selected country, in this case, Turkey. The main use of this plot is for users to observe the overall level of similarity and any obvious outliers, assessing the appropriateness of the selected search space or similar measure.

A problem in the output space has risen due to the window approach in the generation of time-series. It is observed that as the period increases, the output space is dominated by non-unique solutions which are the shifted versions of the time-series. This behavior is more significant when DTW and LCS due to their ability to warp time by repeating and skipping points respectively - assigning same score for shifted versions. To solve this problem, another algorithm is written on top of the sorting algorithm. It is decided that a solution is considered non-unique if the dates for the same countries are in the range of one period. This ensures that even if multiple time-series from the same country are chosen, the series used for prediction do not have any common data points.

4.2 Time-Series Analysis

The time series search produces an *initial set of reference time series*, each of which contain a pattern matched with the time window of the target region. Occasionally, a time series may contain two or more matched patterns. VASAP will list the same time series multiple times, one for each matched pattern.

A time series with the best matching pattern does not mean that the time series is most suitable for predicting the trend of the target region. There are many other factors that affect the suitability of a time series. For example, (i) the pattern may feature some anomalies as observed in the visualization of matched patterns (e.g., Ireland in Fig. 3); (ii) there may not be a sufficient amount of data succeeding the matched pattern. So the time series cannot offer much data to aid the prediction; (iii) the pattern may happen in a time period when a different variant was spreading or has very different intervention policies; (iv) the pattern may happen in a region that features very different geographical, demographic, social, and other characteristics and such difference is significant enough to undermine the suitability of the time series to be included in the reference set.

In general, the algorithm for identifying matching patterns in time series does not have adequate knowledge to decide if a time series in the initial reference set should be removed from the set. When we consulted some domain experts in LMICs about this work, we found that most of these experts were hugely knowledgeable about the COVID-19 situations and intervention policies in their neighboring countries. It is thus much more appropriate and effective to let domain experts decide as to whether a time series should be included or excluded in the reference set, as long as the visualization is suitably designed to support their tasks.

User-led Refinement. As show in Fig. 4, VASAP provides an individual visualization plot for each time series in the reference set as well as the time series of the target region. All these time series are displayed with their own original time axis, and the time window of the target region and all matched patterns are highlighted, so that the users can determine if there is enough data following the matched pattern, or if the matched pattern took place in a period featuring a different variant.

Each time series shown in Fig. 4 is associated with a switch button for a user to exclude it from the reference set (the default is inclusion). It highlights a few example time series that a user may consider unsuitable for predicting the trend of Turkey for the period after April 1, 2022 and decide to exclude from the reference set. For example, a user may decide to exclude Ireland partly because of the anomaly , and partly because of the geographical and demographic differences of Turkey and Ireland. A user may decide to exclude Hungary because the matched pattern is in a different phase; and exclude Singapore because of the geographical, demographic, and social differences of Turkey and Singapore. In some cases, users may make an inclusion or exclusion decision based on their knowledge. In other cases, users may wish to visualize additional data in order to corroborate an inclusion or exclusion decision. For the latter, VASAP enables users to visualize additional data.

In addition to the aforementioned two categories of data (cases and fatalities), the categories of data available for extra data visualization also includes the current number of ICU parities, number of people being vaccinated, number of people being tested, and stringency index. Most of these categories of data also have different types, such as daily, weekly cumulative, raw, and normalized. For example, Bulgaria in Fig. 4 has low vaccination and high death

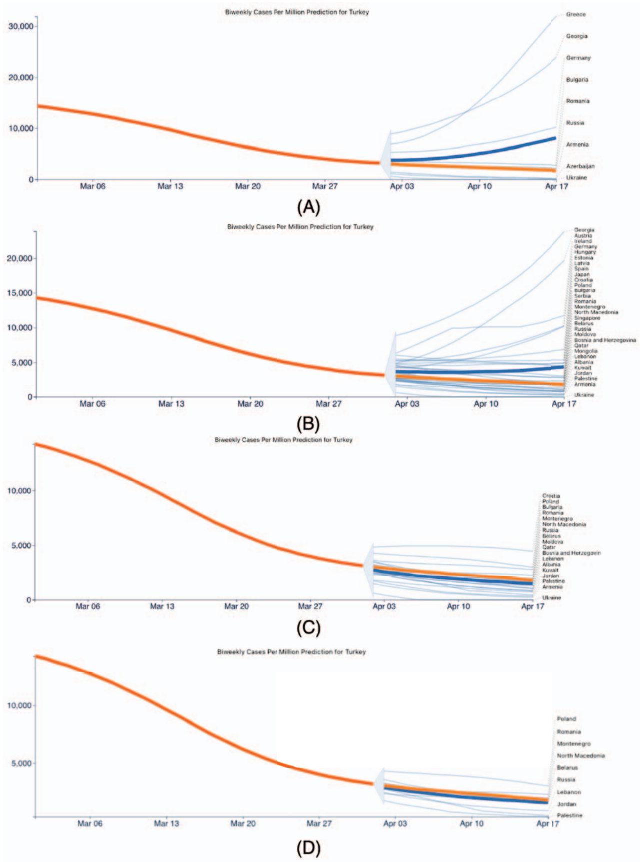

Figure 5: The baseline data from all reference time series is aligned together to generate a prediction plot, where the mean of the baseline data is highlighted as the suggested prediction while all baseline data convey the uncertainty of the prediction. Because we now have the actual data of Turkey after April 1, 2022, we can compare the four predictions made in different steps, i.e., (a) using the 9 countries that might be assumed without using VASAP; (b) using the top 30 countries discovered by time series search; (c) using the 18 countries after a user excluded 12 countries when viewing multiple line charts in Fig. 4; and (d) using the remaining 9 countries after a user excluded 9 countries when viewing comprehensive country dashboard.

values, a user may decide to exclude Bulgaria from the reference set.

Because the main task at this step is for users to filter out unsuitable time series from the reference set, the plots in Fig. 4 enable users to focus on each time series individually without less distraction (as it would be with Fig. 1). The individual plots also serve as external memorization, making it easier for a user to revisit an inclusion/exclusion decision and change one's mind if necessary. The visualization at this stage empowers users to use their domain expertise about different factors (e.g., COVID-19 variants, intervention policies, anomalies in time series data, geographical, demographic, social, and other characteristics of different regions, and so on) to analyze the data and finalize the reference set.

4.3 Time-Series Prediction

Once the reference set is finalized, each time series in the reference set provides a piece of baseline data for prediction. The baseline data is the data points succeeding the matched pattern in the time series. The baseline data from all time series in the reference set is amalgamated together into a single aligned bundle shown light blue in Fig. 5, where the mean of the baseline data is shown in dark blue. In some situations, the time window of the target region is not set at the end of the time series (i.e., the most recent period). Hence, the data succeeding the time window can be considered as posterior validation data. Such data is shown as an orange line segment.

The mean prediction line is a suggested prediction, while the bundle of baseline data convey the uncertainty. The prediction can be made with two categories of data (cases and fatalities).

Fig. 5 compares the four predictions that were made (a) based on assumption only, (b) search only, (c) observation and analysis without viewing extra data, and (d) further analysis in conjunction with visualizing extra data. Because we have now the actual data of Turkey after April 1, 2022, we can observe the similarity between the mean prediction in dark blue and the actual data in orange. We can observe the gradual improvement of the prediction from (a) to (d), demonstrating the benefit of the search-analysis-prediction workflow supported by VASAP. Currently only one aggregate visualization is available in the system but other configurations such as highlighting worst and best case curves, sorting curves by their geographical similarity can be considered as upgrades.

5 Evaluation

After the initial deployment of VASAP, we evaluated it with the domain experts to validate the usefulness and usability of the current prototype, as well as making continuous adjustments and improvements in an agile manner. The evaluation went through the following three stages: 1) an initial expert consultation, 2) the engagement survey, and 3) the think-aloud meeting. In this section, we discuss the methodology and results of each stage.

Stage 1: Initial Expert Consultation

We conducted an initial expert consultation on the tool on 9 December 2021 with two domain experts from Malaysia. After a brief introduction, we carried out an in-depth discussion of the potential usage of the tool and its features for an hour. The health experts were enthusiastic about the potential that VASAP could offer and indicated that the search-based approach would address the need for supporting healthcare experts in anticipating the trend of COVID-19 infection patterns in their region. They confirmed that the healthcare experts there were normally well-informed about various demographic factors and situations of COVID-19 in the neighboring countries and regions, and they could make informed-decisions in selecting a subset of counties for comparative analysis with a target region. They also recommended us to include other data streams that are more reliable in underdeveloped regions.

Stage 2: Engagement Survey

Methodology. The engagement survey had two main intentions: 1) to gather reactions to a prediction tool for COVID-19 and 2) to find potential experts for the follow-up evaluation. Given that potential users of VASAP are more likely to be domain experts with a certain background in COVID-19 data analysis, we chose snowball sampling method [10] to recruit participants efficiently.

The survey was conducted from 3 February 2022 through 8 March 2022 via the online survey tool Qualtrics [21]. The questionnaire had three main sections. The participants were first asked to fill in two demographic questions about their country and occupation. Then a series of questions regarding prediction tools for COVID-19 and finally, whether they would be interested in a follow-up user evaluation. After a 34-day questionnaire recruitment phase, we got 48 responses in total.

Results. Out of the 48 responses, 44 were valid. The results of the first two demographic questions of the questionnaire showed that the data we collected were geographically distributed mainly in Malaysia, Singapore, and South Africa, with 66% of respondents in management and professional, and related fields or in government. In the second part of the questionnaire, 45% of people said that they are driven by the need for observing COVID-19 data and making data-informed decisions at work (Fig. 6a); 57% of people said they had used some online COVID-19 data analysis tools (Fig. 6b).

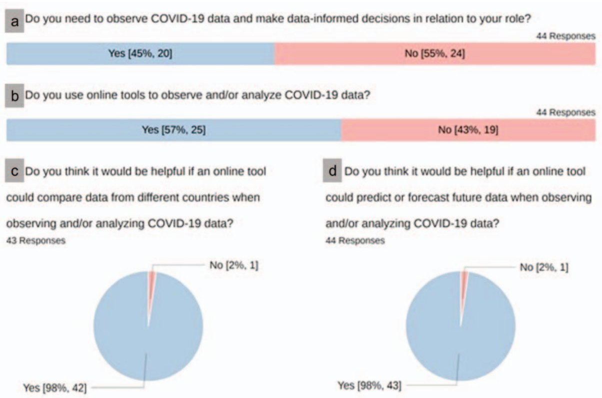

Figure 6: Statistics from the engagement survey.

When asked if they think it would be helpful to have a tool that can compare data with different countries and predict future data when analyzing and / or observing COVID-19 data, 98% of the respondents are positive (Figs. 6c & d). Finally, 28 volunteers indicated their willingness to participate in the future evaluation of VASAP.

STAGE 3: THINK-ALOUD MEETING

METHODOLOGY. A follow-up think-aloud meeting was held with four domain experts recruited through the engagement survey. These sessions lasted 60∼90 minutes per participant and were conducted online. Four experts are all familiar with COVID-19 related data and have experience with the use of data analysis tools.

Procedure: The evaluation was conducted in four stages. We first introduce the objective and setup of the experiment and obtain verbal consent to record the experiment. We then present VASAP via screen-share and demonstrate functionalities through a simple example. The participants are then invited to execute a series of activities and asked to think-aloud when they use the tool. An open discussion on reflections and future directions took place at the end.

Activities: The evaluation process is run through the following four activities: **A1**: *How would you use the tool to search for COVID-19 related data in a country in a certain time period? Do the detected data from other countries produced by the tool are similar to the one you are searching for?* **A2**: *How do you feel about the different similarity detection metrics?* **A3**: *Would you be able to do a comprehensive country comparison, for example: check if the countries returned from the one-dimensional time-period search has similar behavior in other indicators.* **A4**: *How would you use the tool to forecast the future COVID-19 related data in a country? Can the forecast result produced by the tool support your work?*

RESULTS. We analysed our notes gathered through the think-aloud evaluation and discuss the consolidated feedback in the following.

Feedback on the Time Series Search panel: After a brief demonstration, all the participants are able to set the target country, data streams, and time periods based on their needs. The participants found that the resulting data streams from other countries show very similar patterns to the target. Apart from the similarity search result, they appreciated the richness of the data streams: *What I like in this tool is that you have extensive data for visualization, that is really good, . . . It is not easy to collect all the data in one repository and then have the visualization tool to see all those patterns.*

Feedback on choosing similarity measures: For choosing different similarity measures, we find that three of the participants tend to choose the default value, while only one chose a different similarity measure. After the follow-up interviews, we found out that it is difficult for them to discern appropriate similarity measures. They tend to think that the default value is the best. The participant who chooses a value different from the default explained that he has no preference for these measures, but since one of them was often used in his previous work, he chooses it due to familiarity.

Feedback on the Comprehensive Country Comparison panel: During the evaluation, we find that participants can choose data streams from different countries and add them to the list to visualize, but still need our intervention and guidance in some operations. The participants found the panel to compare data streams across countries useful. The third participant appreciated the ability to manually enter country names and customize country lists for comparisons, and stated: *I think it's great. It helps all these university administrators with student academic affairs track information related to international students' COVID-19 status, especially looking at the different countries. I think that would bring great potential, especially at the governance level. It is like a different way to use the comparison panel to see time series and COVID patterns in different countries, and you can also select the countries you want.* Although this is inconsistent with the original intention of the VASAP design, we are glad that users can find their own way of using the tool.

Feedback on the Observation-based Forecasting panel: The participants showed more interested in the Observation-based Forecasting panel compared to the other views. All participants mention that currently they do not have an available tool for future data prediction, it would be a great help to have such a tool.

A participant, who runs a district hospital in South Africa, mentioned that they are in great need of having a suitable tool for short-term prediction of daily / weekly cases. This will help them in the continuous management of hospitals in the context of COVID. Currently, he uses Excel's built-in model to make predictions. As for the prediction results, the participants do not have extremely high requirements for absolute accuracy. They believe that the prediction of the trend is already accurate and very useful.

REFLECTION

Both the engagement survey and think-aloud meeting have confirmed that there is a high demand from domain experts for a COVID-19 data analysis tool that can provide multi-country data comparisons and future data forecasts. The results above provide evidence for the practical relevance of VASAP. The domain experts also put forward several suggestions for the improvement of the UI design and user experience, such as: adding further instructions, and recommendations for distance metrics, and so on.

6 CONCLUSION

In this paper, we introduce a visual analytics approach that facilitates human-guided and data-driven predictions through a search-analyze-predict process. We demonstrate how challenges in developing a data-driven prediction framework could be addressed by incorporating a multi-metric, multi-criteria search algorithm supported by visualizations enabling the comparative analysis and the manual curation of the search results, with an eventual goal to arrive at a data-driven prediction of disease progress for a chosen location. Our evaluation of the proposed approach with domain experts provided evidence that data-driven predictions of the progress of the pandemic in the near future has the potential to be an effective and useful method to support localized, short-term decision making, and we noted that inaccuracies in the predictions are found to be more acceptable in situations such as these, indicating potentially the importance of data-driven techniques as a situation awareness and assessment tool.

As both reported in related literature and re-iterated in our empirical research, such data-driven but human-guided tools for forecasting have the potential to be viable and valuable assets, especially for countries lacking extensive epidemiological modelling capacity. We argue, therefore, that approaches such as the one introduced in this paper could contribute significantly to pandemic preparedness globally, and we call for further research into integrating more sophisticated modelling methods and advanced visual analytics capability in openly available, easy-to-maintain tools.

ACKNOWLEDGMENTS

Authors would like to thank UKRI/EPSRC funding for the project "RAMP VIS: Making Visual Analytics an Integral Part of the Technological Infrastructure for Combating COVID-19" (EP/V054236/1), and all colleagues in the RAMPVIS project and in the Scottish COVID-19 Response Consortium (SCRC).

REFERENCES

[1] Center for Systems Science and Engineering, Johns Hopkins University. Covid-19 dashboard. `https://coronavirus.jhu.edu/map.html`, Last accessed: March 2022.

[2] R. Aguas, L. White, N. Hupert, R. Shretta, W. Pan-Ngum, O. Celhay, A. Moldokmatova, F. Arifi, A. Mirzazadeh, H. Sharifi, K. Adib, M. N. Sahak, C. Franco, and R. Coutinho. Modelling the covid-19 pandemic in context: an international participatory approach. *BMJ Golbal Health*, 5(12):e003126, 2020.

[3] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski. Visualizing time-oriented data—a systematic view. *Computers & Graphics*, 31(3):401–409, 2007. doi: 10.1016/j.cag.2007.01.030

[4] J. Burmeister, J. Bernard, and J. Kohlhammer. Eurovis workshop on visual analytics (eurova). 2021. doi: 10.2312/eurova.20211098

[5] M. Chen, A. Abdul-Rahman, D. Archambault, J. Dykes, A. Slingsby, P. Ritsos, T. Torsney-Weir, C. Turkay, B. Bach, R. Borgo, A. Brett, H. Fang, R. Jianu, S. Khan, S. Laramee, L. Matthews, P. H. Nguyen, R. Reeve, J. C. Roberts, F. Vidal, Q. Wang, J. Wood, and K. Xu. RAMPVIS: Answering the challenges of building visualisation capabilities for large-scale emergency responses. *Epidemics*, 39:ID–100569, 2022.

[6] M. Elshehaly, R. Randell, M. Brehmer, L. McVey, N. Alvarado, C. P. Gale, and R. A. Ruddle. Qualdash: adaptable generation of visualisation dashboards for healthcare quality improvement. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):689–699, 2020.

[7] P. Esling and C. Agon. Time-series data mining. *ACM Computing Surveys (CSUR)*, 45(1):1–34, 2012.

[8] A. Fakhrazari and H. Vakilzadian. A survey on time series data mining. In *2017 IEEE International Conference on Electro Information Technology (EIT)*, pp. 476–481, 2017. doi: 10.1109/EIT.2017.8053409

[9] A. Gogolou, T. Tsandilas, T. Palpanas, and A. Bezerianos. Comparing similarity perception in time series visualizations. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 08 2018. doi: 10.1109/TVCG.2018.2865077

[10] L. A. Goodman. Snowball sampling. *The annals of mathematical statistics*, pp. 148–170, 1961.

[11] GOV.UK. Coronavirus (COVID-19) in the UK. `https://coronavirus.data.gov.uk/`, last assessed in March 2022.

[12] W. Jentner, F. Sperrle, D. Seebacher, M. Kraus, R. Sevastjanova, M. T. Fischer, U. Schlegel, D. Streeb, M. Miller, T. Spinner, E. Cakmak, M. Sharinghousen, P. Meschenmoser, J. Görtler, O. Deussen, F. Stoffel, H.-J. Kabitz, D. A. Keim, M. El-Assady, and J. F. Buchmüller. Visualisierung der covid-19-inzidenzen und behandlungskapazitäten mit coronavis. In A. Karsten and S. Voßschmidt, eds., *Resilienz und Pandemie : Handlungsempfehlungen anhand von Erfahrungen mit COVID-19*, pp. 176–189. Kohlhammer, Stuttgart, 2022.

[13] E. Keogh, S. Chu, D. Hart, and M. Pazzani. *SEGMENTING TIME SERIES: A SURVEY AND NOVEL APPROACH*, pp. 1–21. doi: 10.1142/9789812565402_0001

[14] S. Khan, P. Nguyen, A. Abdul-Rahman, E. Freeman, C. Turkay, and M. Chen. Rapid development of a data visualization service in an emergency response. *IEEE Transactions on Services Computing*, 15(2):1251–1264, 2022.

[15] S. Khan, P. H. Nguyen, A. Abdul-Rahman, B. Bach, M. Chen, E. Freeman, and C. Turkay. Propagating visual designs to numerous plots and dashboards. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):86–95, 2022.

[16] R. Kincaid. Signallens: Focus+ context applied to electronic time series. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):900–907, 2010.

[17] E. S. McBryde, M. T. Meehan, O. A. Adegboye, A. I. Adekunle, J. M. Caldwell, A. Pak, D. P. Rojas, B. M. Williams, and J. M. Trauer. Role of modelling in covid-19 policy development. *Paediatric respiratory reviews*, 35:57–60, 2020.

[18] L. Padilla, H. Hosseinpour, R. Fygenson, J. Howell, R. Chunara, and E. Bertini. Impact of covid-19 forecast visualizations on pandemic risk perceptions. *Scientific Reports*, 12(1), Feb. 2022. doi: 10.1038/s41598-022-05353-1

[19] L. A. Post, J. S. Lin, C. B. Moss, R. L. Murphy, M. G. Ison, C. J. Achenbach, D. Resnick, L. N. Singh, J. White, M. J. Boctor, S. B. Welch, and J. F. Oehmke. SARS-CoV-2 wave two surveillance in East Asia and the Pacific: Longitudinal trend analysis. *Journal of Medical Internet Research*, 23(2):e25454, 2021.

[20] B. Preim and K. Lawonn. A survey of visual analytics for public health. *Computer Graphics Forum*, 39(1):543–580, 2020.

[21] Qualtrics. Online survey software: Our software powers +1b surveys every year. `https://www.qualtrics.com`, last assessed in March 2022.

[22] M. Roser. Our world in data. `https://ourworldindata.org/coronavirus`, last assessed in March 2022.

[23] Y. Wang, F. Han, L. Zhu, O. Deussen, and B. Chen. Line graph or scatter plot? automatic selection of methods for visualizing trends in time series. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1141–1154, 2018. doi: 10.1109/TVCG.2017.2653106

[24] Y. Zhang, Y. Sun, L. Padilla, S. Barua, E. Bertini, and A. G. Parker. Mapping the landscape of COVID-19 crisis visualizations. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, may 2021. doi: 10.1145/3411764.3445381