EnsembleVis: A Retrospective and Early History

Qiru Wang*
University of Nottingham

Robert S. Laramee†
University of Nottingham

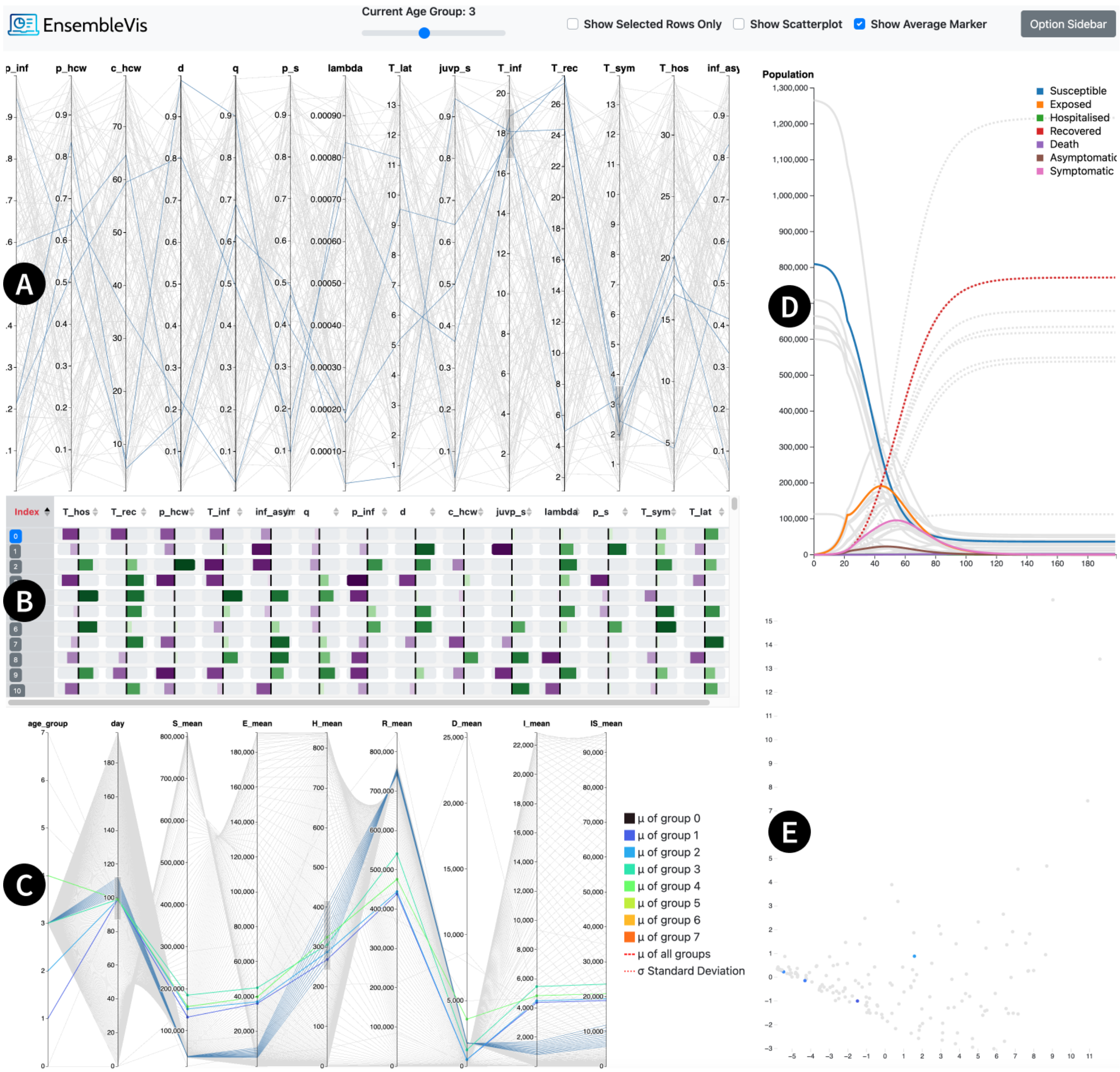

Figure 1: An overview of EnsembleVis, an interactive dashboard designed for visualizing the input parameters and outcomes of an ABC-smc inference model used for analyzing COVID-19 data collected during the first wave of the outbreak in Scotland [1]. Its main purpose is to facilitate a better understanding of the complex dynamics of the pandemic by presenting the relationships between different sets of input parameters and the resulting outcomes in a clear and intuitive manner. A detailed description of the dashboard can be found in Section 4.

Abstract

Index Terms: Visual Analytics—Information Visualization—Emergency Response—Visual Design

1 Introduction and Motivation

The Scottish COVID-19 Response Consortium (SCRC) [3], in collaboration with the Royal Society's call to action in March 2020, has taken a proactive approach to address the need for enhanced epidemi-

*e-mail: qiru.wang1@nottingham.ac.uk
†e-mail: robert.laramee@nottingham.ac.uk

ological models of COVID-19 transmission. This joint effort, known as Rapid Assistance in Modelling the Pandemic (RAMP) [2], aims to foster a deeper comprehension of the consequences associated with various exit strategies from lockdown measures. Moreover, this consortium has attracted the involvement of distinguished scientists and experts from diverse organizations both within the United Kingdom and abroad, thus augmenting the collective knowledge base and ensuring comprehensive expertise in specialized domains.

RAMPVis [4] is a group of volunteers specialized in Data Visualization and Visual Analytics (abbreviated as VIS). This group has willingly come forward to lend its specialized skills and knowledge in order to provide invaluable support to the consortium's modelers and epidemiologists.

Serving as the volunteer team responsible for providing visualization support to one of the six epidemiological models developed by the SCRC modelers [8], our main objective is to provide VIS researchers and practitioners with valuable insights gleaned from our research and development (R&D) activities conducted during the COVID-19 pandemic. In an effort to predict the potential impact of diverse interventions, both modelers and epidemiologists have actively utilized COVID-19 data, employing a method known as Uncertainty Quantification (UQ). This process seeks to measure uncertainties through the application of mathematical models and simulations. However, the modelers are confronted with significant challenges, including the aspects of expert elicitation and effective communication. In other words, there is a need for software engineering effort coupled with visualization to provide support for the validation and verification tests, and to create efficient workflows between the modelers and the epidemiologists [6].

In addressing these hurdles, Data Visualization and Visual Analytics (VIS) emerges as a potent tool, offering the capacity to significantly enhance and streamline their collaborative workflows [24]. While our work may not have showcased the state-of-the-art VIS techniques, it effectively delivered rapid and practical VIS support to the modelers during an exceptional and demanding time. The exemplification of a fully-virtual collaboration among researchers from various UK institutions epitomizes the spirit of interdisciplinary cooperation in the fight against the pandemic.

2 Background and Related Work

VIS has been widely utilized in critical applications such as emergency responses and healthcare, assisting public officials and decision-makers in understanding intricate datasets and extracting useful, actionable insights from them [10]. VIS has also played a prominent role in disseminating COVID-19 information through various media channels, it has significantly contributed to enabling more efficient and clearer public communication, facilitating a broader understanding of the crisis [16].

In our work, our primary objective was to extend support through VIS to two distinct user groups. Firstly, the modelers, who could significantly benefit from VIS in comprehending their models more effectively and fine-tuning them accordingly. Secondly, to the epidemiologists, whom VIS could assist in interpreting the outcomes of these computational models. Our work is included in multiple publications [8,11,17,18,23], where it functioned as the preliminary VIS prototype, shaping a portion of their respective studies.

VIS for Emergency Response

Previously as co-authors, we have detailed the related work focusing the use of VIS in emergency response, refer to the related work section in Chen et al. [8]. The aforementioned literature review laid the foundation and was conducted prior to the development of our current study in 2020.

Maciejewski et al. [20] develop a VIS toolkit for analyzing the effect of decision measures enforced during a simulated pandemic, the tool was later utilized by Indiana State Department of Health during an H1N1 (swine flu) outbreak. Ribicic et al. [22] leverage VIS with the intention of delivering real-time feedback derived from flood simulations to non-expert users, while Konev et al. [19] use VIS to support decision-making in flooding scenarios.

Jeitler et al. [15] use VIS to analyze social media data to aid rescue teams, specifically in terms of optimally allocating resources during emergency response situations. Similarly, Nguyen and Dang [21] harness social media data, paired with VIS, to facilitate and enhance post-earthquake resource allocation and rescue effort.

In contrast to the majority of previous studies mentioned here that typically focus on preparing for future emergencies, our work was undertaken during the COVID-19 pandemic as a rapid response to an ongoing emergency.

VIS for COVID-19 Data Modeling

In the rest of the section, we focus on the use of VIS in aiding the computational modeling of COVID-19 data. These studies were not published or available to us during the development of our work. In fact the use of VIS in epidemiological modeling was rare, both modelers and epidemiologists might be unaware that they had such a potent instrument readily available [8].

He et al. [13] developed an SEIR (Susceptible, Exposed, Infected, and Recovered) model for spread prediction by leveraging COVID-19 data obtained from the Hubei province in China. They employed a variety of 2D plots for estimating the parameters of the model and interpreting the outcomes yielded by the model. Godio et al. [12] took the same approach in their development of an SEIR model for the Lombardy region in Italy.

The IHME COVID-19 Forecasting Team [14] took the application of data visualization (VIS) a step further in their development of SEIR model for accessing social distance mandates, they extend the use of VIS to include choropleth and violin plots, and small multiples for 2D plots.

Chinazzi et al. [9] developed a model for simulating the effectiveness of international travel restrictions in containing the spread of COVID-19. Besides the employment of 2D plots for refining their models, they also utilized a range of geospatial visualizations. This allowed them to more effectively interpret the results generated by their models. The use of geospatial visualizations is also adopted by Alvarez Castro and Ford [7] in their development of a model for analyzing the transmission in a UK university campus.

Contrary to these studies that showcase the efficacy of VIS in supporting computational modeling of COVID-19 data with a primary focus on the model's development, as they are formulated by the modelers, our study takes a different approach. We focus our attention on refining VIS as a potent tool that can significantly enhance the computational modeling of COVID-19 data, all viewed through the unique lens of a VIS practitioner.

3 Data Description

Table 1: 16 input parameters used for the ABC-SMC inference model.

ID	Compart	Description
S	0	Number of susceptible individuals (not infected).
E	1	Number of infected individuals but not yet infectious (exposed).
E_t	2	Number of exposed individuals and tested positive.
I_p	3	Number of infected and infectious symptomatic individuals but at pre-clinical stage (show yet no symptoms).
I_t	4	Number of tested positive individuals that are infectious.
I1	5	Number of infected and infectious asymptomatic individuals: first stage.
I2	6	Number of infected and infectious asymptomatic individuals: second stage.
I3	7	Number of infected and infectious asymptomatic individuals: third stage.
I4	8	Number of infected and infectious asymptomatic individuals: last stage.
I_s1	9	Number of infected and infectious symptomatic individuals: first stage.
I_s2	10	Number of infected and infectious symptomatic individuals: second stage.
I_s3	11	Number of infected and infectious symptomatic individuals: third stage.
I_s4	12	Number of infected and infectious symptomatic individuals: last stage.
H	13	Number of infected individuals that are hospitalized.
R	14	Number of infected individuals that have recovered from the infection.
D	15	Number of deceased individuals due to the disease.

Table 2: Description of output parameters.

Name	Description
iterID	The simulation number.
age_group	The age group of the population.
comparts	The epidemiological compartment number or the parameter space (See Table 1).
value	The population of that compartment at the end of the simulation run.

The data used in our work includes simulation parameters and outcomes from an ABC-SMC inference model [25] built by a group of modelers from University of Edinburgh, University of Exeter, University of Glasgow, and London School of Hygiene & Tropical Medicine. The underlying pandemic data was collected during the first wave of the outbreak in Scotland. The model was built to analyze the data and infer the parameters of the model that best fit the data. The model was run 1,000 times with different random seeds to generate a set of 160 parameter sets. The model has 16 input parameters (See Table 1) and 4 output parameters (See Table 2). The resulting output file is a CSV file that is over 6GB in size.

It is worth mentioning that after plotting the output data using a line chart, an error was immediately spotted, see Figure 1D, where an unusual spike at day 20 can be observed. The modelers were notified and the bug was fixed. However, the rectified output file was never made available to us.

4 EnsembleVis

This section presents the experience behind our fully virtual collaboration between researchers from multiple UK institutions, chronicling the development of EnsembleVis.

First Meeting - 27 July 2020

On 27 July 2020, amid the UK's first national lockdown and stricter measures imposed by local authorities, we convened the initial virtual meeting with VIS researchers from King's College London, Loughborough University, Swansea University, University of Nottingham, University of Warwick, and University of Oxford.

During the meeting, we received an overview of the SCRC and the responsibilities of the visualization volunteer team. Our assigned task was to create visualizations for the model, with the purpose of allowing the modelers to analyze the outcomes of the model.

Following the initial meeting, we engaged in email correspondence with the modelers to delve into the visualization requirements. The modelers shared a comprehensive list of parameters and model outcomes, along with the corresponding outcome data.

First Commit - 14 Sep 2020

We proceeded to create an initial prototype of the visualization, which was subsequently reviewed by the modelers. Incorporating their input, we refined the prototype during our weekly internal discussions. On 14 Sep 2020, England introduced the 'rule of six', which banned any gatherings above six. On the same day, we made our first commit to a GitHub repository, signifying the commencement of our development. At the same time, we began preprocessing the data.

A week after the initial commit, the UK witnessed the implementation of additional restrictions, such as mandatory work from home and a 10pm curfew.

First Visualization, Second Meeting - 5 Nov 2020

On 5 Nov 2020, the first day of the second national lockdown in the UK, we completed the first visualization, a parallel coordinates plot, see Figure 2. This followed by the second meeting with VIS researchers from other institutions, where we received feedback on the visualization, on 6 Nov 2020.

Third Meeting - 11 Nov 2020

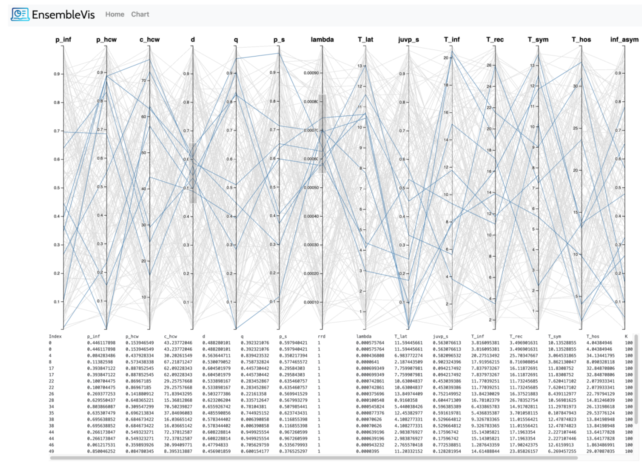

Figure 2: The first visualization, a parallel coordinates plot visualizing all 160 configurations of the model, was completed on 5 Nov 2020. This visualization was developed utilizing D3.js and incorporated a brushing feature, enabling users to filter the configurations.

On 11 Nov 2020, the group convened for the third meeting, where we received further feedback on the visualization. As per the modelers' requests conveyed through emails, we incorporated a line chart to depict the model outcomes, see Figure 3.

Fourth Meeting - 25 Nov 2020

On 25 Nov 2020, the group convened for the fourth meeting, held just a day after the announcement of the gathering rules for Christmas in the UK. During the meeting, we received feedback on the new visualization, a table with glyphs, see Figure 1B. We incorporated this table view featuring glyphs to visualize all 160 input parameter configurations, following discussions with the modelers. Each parameter is symbolized by a glyph, with the color of the glyph corresponding to its deviation from the average value. The view provides the functionality to sort the parameters according to their values and can be dynamically updated by brushing the parallel coordinates plot for the input parameters (Figure 1A).

Fifth Meeting - 9 Dec 2020

On 9 Dec 2020, a week after the end of the second national lockdown in the UK, with England facing a stricter three-tier restriction policy, the group convened for the fifth meeting. At this point, we still had not met with the modelers, all communications and discussions took place through emails.

Sixth Meeting - 10 Dec 2020

On 10 Dec 2020, we finally met with modelers from the University of Edinburgh, the University of Exeter, the University of Glasgow, The London School of Hygiene & Tropical Medicine, for the first time. In contrast to sharing screenshots via email and deploying a website with a live view of our development (which they might not have been proficient in using), we delivered a live presentation, fielding numerous questions. The modelers were extremely pleased with the visualization, and a list of feedback was provided:

1. The modelers found the parallel coordinates plot very useful, and requested the incorporation of another one for the model outcomes. We implemented this as shown in Figure 1C.

2. The modelers requested all the simulation results to be displayed in Figure 3, with the current one being highlighted. We implemented this as shown in Figure 1D.

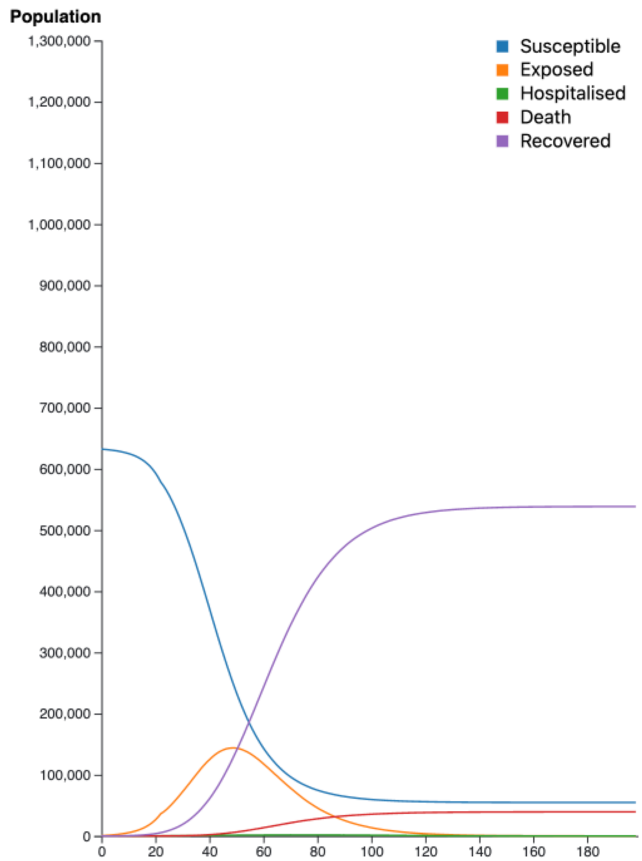

Figure 3: A line chart was added to the visualization on 11 Nov 2020, to visualize the model outcomes. The x-axis of the chart corresponds to the number of days since the first available date in the Scottish dataset (specific date unknown to us), while the y-axis represents the population. To differentiate between different population categories, a color map was employed: susceptible (blue), exposed (orange), death (red), hospitalized (green), and recovered (purple). The final version of this chart which includes all the simulation results, is shown in Figure 1D.

3. The modelers requested the incorporation of a scatterplot to visualize the model outcomes, specifically, a Principal Component Analysis (PCA) result obtained from other volunteers. We implemented this as shown in Figure 1E.

Furthermore, we received the exciting news that funding had been successfully secured, leading to the transition of our voluntary work to a team of paid developers, who would continue with further development of the project on a future date.

Last Commit - 28 Apr 2021

By 28 Apr 2021, the UK began a gradual easing of measures, although the prohibition on mixing between households was still in effect. On this day, we made our last commit to the GitHub repository, this act signified the completion of our work, as we had smoothly transitioned all tasks to a team of paid developers. The final version of our work is shown in Figure 1.

During the entire development process, our meetings were exclusively conducted virtually, and our communication relied heavily on email correspondence. Despite the lack of in-person interactions, we successfully met the modelers' requirements and delivered a highly satisfactory VIS solution.

5 Limitations

Due to the impact of the pandemic, the project was conducted in a fully virtual manner, with all meetings and discussions taking place online, between a very large group of researchers from different disciplines. This has resulted in a number of limitations, which we will discuss in this section.

5.1 Lack of Novel and Advanced Visual Designs

Operating under a time constraint, the primary objective of our project centered on offering immediate visualization assistance to the modelers. Thus, we were unable to explore the inclusion of innovative and advanced visual design approaches. Instead, we integrated a series of classic visualizations, such as line charts and scatterplots. These are visual elements commonly leveraged by modelers and epidemiologists in their day-to-day research. Interestingly, the modelers welcomed the introduction of a less conventional (to them) visualization technique: parallel coordinates. They had never before employed this visual design, and its introduction proved beneficial to their research. Consequently, they expressed a desire for the incorporation of an additional parallel coordinates to assist in the visualization of model outcomes.

We believe that this is a testament to the effectiveness of advanced visual designs in enhancing the modelers' understanding of their models, this signals the possibility for future inclusion of more sophisticated visual designs.

5.2 Lack of Proper Requirement Gathering

We were unable to meet with the modelers and epidemiologists until the very last meeting. Instead, we had to rely on email correspondence, which was arguably not as effective as face-to-face or even virtual meetings.

In a traditional software engineering project, the requirements are gathered through a series of meetings and discussions with the end users. This did not happen in our case.

This resulted in a lack of proper requirement gathering, which in turn led to a number of issues during the development process. For example, the modelers made ad-hoc requests to incorporate different visualizations at different stages of the project, resulting in unexpected changes on the development side. This could have been avoided if we had a better understanding of their requirements from the beginning.

5.3 Dynamic Group Membership

The group membership was dynamic, with researchers joining and leaving the group at different stages of the project. This resulted in a lack of continuity, as the newcomers had to spend time to familiarize themselves with the project. Furthermore, members came from different disciplines, with different levels of expertise in visualization. This has resulted in a lack of consistency in the development process, as different members have different ideas on how to implement the visualization. The responsibility of each member, apart from the only developer in the group, was not clearly defined.

5.4 Unclarified Project Direction

The exact direction of the project was not clearly defined from the beginning. Many details remained unknown to us during the development process, such as the exact purpose of the visualization, the target audience, and the end product. Consequently, the final product suffered from a non-ideal utilization of screen-space, as more visualizations were requested to be added, the implementation of a multi-screen display design or collapsible views became time-constrained and unachievable.

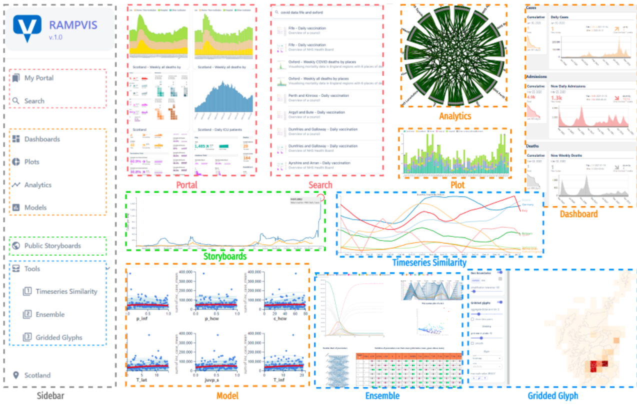

Figure 4: EnsembleVis has undergone extensive development by multiple UK institutions and is currently maintained by the Oxford e-Research Centre at the University of Oxford, serving as a vital element of the RAMPVIS infrastructure. It is well-prepared to offer rapid and invaluable visualization support for future emergency responses. Image courtesy of Rydow et al. [23].

6 Conclusions

In this paper, we have presented the stories behind the development of EnsembleVis, an interactive dashboard designed for visualizing the input parameters and outcomes of an ABC-smc inference model used for analyzing COVID-19 data collected during the first wave of the outbreak in Scotland. Our voluntary work has contributed to various publications [8, 11, 17, 18, 23].

Given the multitude of uncertainties and challenges during this exceptional period, a considerable amount of information was unavailable to us during the development process. It was only through the Scottish COVID-19 Response Consortium Stakeholder Report [5], published in late 2021, and various publications that unveiled the remarkable endeavors undertaken by other volunteer teams, that we gained additional insights and details.

We hope that our experience serves as a valuable source of insights on how VIS research and techniques can play a crucial role in emergency response initiatives and aid in effectively preparing for future emergencies.

References

[1] Covid19_EERAModel. SCRC, Sept. 2020.

[2] Rapid Assistance in Modelling the Pandemic: RAMP — Royal Society, 2020.

[3] University of Glasgow - The Scottish COVID-19 Response Consortium, 2020.

[4] Visualization and Visual Analytics in Support of Rapid Assistance in Modelling the Pandemic (RAMP), 2020.

[5] Y. Abdalla, H. Auty, L. Boden, A. Brett, M. Chen, R. Dundas, L. Matthews, I. McKendrick, D. Mellor, and R. Reeve. Scottish COVID-19 Response Consortium Stakeholder Report. Technical report, 2021.

[6] G. J. Ackland, J. Panovska-Griffiths, W. Waites, and M. E. Cates. The Royal Society RAMP modelling initiative. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380(2233):20210316, Aug. 2022. doi: 10.1098/rsta.2021.0316

[7] D. Alvarez Castro and A. Ford. 3D Agent-Based Model of Pedestrian Movements for Simulating COVID-19 Transmission in University Students. *ISPRS International Journal of Geo-Information*, 10(8):509, Aug. 2021. doi: 10.3390/ijgi10080509

[8] M. Chen, A. Abdul-Rahman, D. Archambault, J. Dykes, P. Ritsos, A. Slingsby, T. Torsney-Weir, C. Turkay, B. Bach, R. Borgo, A. Brett, H. Fang, R. Jianu, S. Khan, R. Laramee, L. Matthews, P. Nguyen, R. Reeve, J. Roberts, F. Vidal, Q. Wang, J. Wood, and K. Xu. RAMPVIS: Answering the challenges of building visualisation capabilities for large-scale emergency responses. *Epidemics*, 39:100569, June 2022. doi: 10.1016/j.epidem.2022.100569

[9] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489):395–400, Apr. 2020. doi: 10.1126/science.aba9757

[10] F. Dusse, P. S. Júnior, A. T. Alves, R. Novais, V. Vieira, and M. Mendonça. Information visualization for emergency management: A systematic mapping study. *Expert Systems with Applications*, 45:424–437, Mar. 2016. doi: 10.1016/j.eswa.2015.10.007

[11] J. Dykes, A. Abdul-Rahman, D. Archambault, B. Bach, R. Borgo, M. Chen, J. Enright, H. Fang, E. E. Firat, E. Freeman, T. Gönen, C. Harris, R. Jianu, N. W. John, S. Khan, A. Lahiff, R. S. Laramee, L. Matthews, S. Mohr, P. H. Nguyen, A. A. M. Rahat, R. Reeve, P. D. Ritsos, J. C. Roberts, A. Slingsby, B. Swallow, T. Torsney-Weir, C. Turkay, R. Turner, F. P. Vidal, Q. Wang, J. Wood, and K. Xu. Visualization for epidemiological modelling: Challenges, solutions, reflections and recommendations. *Phil. Trans. R. Soc. A.*, 380(2233):20210299, Oct. 2022. doi: 10.1098/rsta.2021.0299

[12] A. Godio, F. Pace, and A. Vergnano. SEIR Modeling of the Italian Epidemic of SARS-CoV-2 Using Computational Swarm Intelligence. *International Journal of Environmental Research and Public Health*, 17(10):3535, Jan. 2020. doi: 10.3390/ijerph17103535

[13] S. He, Y. Peng, and K. Sun. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dyn*, 101(3):1667–1680, Aug. 2020. doi: 10.1007/s11071-020-05743-y

[14] IHME COVID-19 Forecasting Team. Modeling COVID-19 scenarios for the United States. *Nat Med*, 27(1):94–105, Jan. 2021. doi: 10.1038/s41591-020-1132-9

[15] A. Jeitler, A. Türkoglu, D. Makarov, T. Jockers, J. Buchmüller, U. Schlegel, and D. A. Keim. RescueMark: Visual Analytics of Social Media Data for Guiding Emergency Response in Disaster Situations: Award for Skillful Integration of Language Model. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 120–121, Oct. 2019. doi: 10.1109/VAST47406.2019.8986898

[16] Johns Hopkins University. COVID-19 Map.

[17] 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, Jan. 2022. doi: 10.1109/TVCG.2021.3114828

[18] S. Khan, P. H. Nguyen, A. Abdul-Rahman, E. Freeman, C. Turkay, and M. Chen. Rapid Development of a Data Visualization Service in an Emergency Response. *IEEE Trans. Serv. Comput.*, 15(3):1251–1264, May 2022. doi: 10.1109/TSC.2022.3164146

[19] A. Konev, J. Waser, B. Sadransky, D. Cornel, R. A. Perdigão, Z. Horváth, and M. E. Gröller. Run Watchers: Automatic Simulation-Based Decision Support in Flood Management. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1873–1882, Dec. 2014. doi: 10.1109/TVCG.2014.2346930

[20] R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller, and S. W. Sanders. A pandemic influenza modeling and visualization tool. *Journal of Visual Languages & Computing*, 22(4):268–278, Aug. 2011. doi: 10.1016/j.jvlc.2011.04.002

[21] H. N. Nguyen and T. Dang. EQSA: Earthquake Situational Analytics from Social Media. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 142–143, Oct. 2019. doi: 10.1109/VAST47406.2019.8986947

[22] H. Ribicic, J. Waser, R. Gurbat, B. Sadransky, and M. E. Groller. Sketching Uncertainty into Simulations. *IEEE Trans. Visual. Comput. Graphics*, 18(12):2255–2264, Dec. 2012. doi: 10.1109/TVCG.2012.261

[23] E. Rydow, T. Gönen, A. Kachkaev, and S. Khan. RAMPVIS: A visualization and visual analytics infrastructure for COVID-19 data. *SoftwareX*, 0(0), May 2023. doi: 10.1016/j.softx.2023.101416

[24] B. Swallow, P. Birrell, J. Blake, M. Burgman, P. Challenor, L. E. Coffeng, P. Dawid, D. De Angelis, M. Goldstein, V. Hemming, G. Mar-

ion, T. J. McKinley, C. E. Overton, J. Panovska-Griffiths, L. Pellis, W. Probert, K. Shea, D. Villela, and I. Vernon. Challenges in estimation, uncertainty quantification and elicitation for pandemic modelling. *Epidemics*, 38:100547, Mar. 2022. doi: 10.1016/j.epidem.2022.100547

[25] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31):187–202, July 2008. doi: 10.1098/rsif.2008.0172