Technological Forecasting & Social Change 166 (2021) 120602

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Forecasting COVID-19 pandemic: Unknown unknowns and predictive monitoring

Jianxi Luo

Data-Driven Innovation Lab, Singapore University of Technology & Design (SUTD), 8 Somapah Road, 487372, Singapore

ARTICLE INFO

Keywords:
COVID-19 pandemic
Uncertainty
Forecasting
Prediction
Monitoring

ABSTRACT

During the current COVID-19 pandemic, there have been many efforts to forecast infection cases, deaths, and courses of development, using a variety of mechanistic, statistical, or time-series models. Some forecasts have influenced policies in some countries. However, forecasting future developments in the pandemic is fundamentally challenged by the innate uncertainty rooted in many "unknown unknowns," not just about the contagious virus itself but also about the intertwined human, social, and political factors, which co-evolve and keep the future of the pandemic open-ended. These unknown unknowns make the accuracy-oriented forecasting misleading. To address the extreme uncertainty of the pandemic, a heuristic approach and exploratory mindset is needed. Herein, grounded on our own COVID-19 forecasting experiences, I propose and advocate the "*predictive monitoring*" paradigm, which synthesizes prediction and monitoring, to make government policies, organization planning, and individual mentality heuristically future-informed despite the extreme uncertainty.

1. Introduction

Since the outbreak of COVID-19, researchers around the world have adopted or developed various data-driven models to predict health indicators or forecast developments and trends in the pandemic in different countries or regions. Noticeable efforts include the forecasts by the Institute of Health Metrics and Evaluation (IHME) at the University of Washington and the MRC center for Global Infectious Disease Analysis at Imperial College London, among others. Table 1 is a list of publicly accessible COVID-19 forecasting programs. The forecasts have focused on future deaths, infections and hospitalizations (IHME, 2020a, 2020b; Woody et al., 2020), peak and ending dates (Yang et al., 2020; Ferguson et al., 2020; Petropoulos and Makridakis, 2020), and the impact of social distancing, travel restrictions, lockdown and reopening, mitigation, and suppression strategies (Kissler et al., 2020; Chinazzi et al., 2020).

Some published studies have attempted to validate the accuracy of specific prediction methods (IHME, 2020a; 2020b; Woody et al., 2020; Yang et al., 2020). However, even the most cited forecasting method, from the IHME, has been found to have model design issues (Woody et al., 2020; Jewell et al., 2020); 70 percent of time, the number of actual deaths fell outside the next-day predictions' 95 percent confidence intervals (Marchant et al., 2020). The IHME team later revised the model (IHME, 2020b), but prediction errors remain high. COVID-19 forecasting appears to be a case where George Box's famous aphorism "*all models are wrong, but some are useful*" applies (Box, 1976). Despite the intrinsic complexity, ambiguity, and uncertainty of such forecasts, some of them have influenced policies or informed policy makers (Grey and MacAskill, 2020; Resnick, 2020). Researchers are learning and improving models and methods on the go, in order to make more accurate and useful predictions (Marchant et al., 2020; Ray et al., 2020).

2. Fundamental challenges to COVID-19 forecasting

The uncertainty that we face for work and life during the pandemic makes data-driven predictions desirable and much needed. However, this uncertainty also makes pandemic prediction difficult. The fundamental challenge is rooted in the nature of the pandemic as a classic "wicked problem" (Rittel and Webber, 1973). Wicked problems are those that are novel, unique, complex, and evolving, with incomplete, contradictory, and changing requirements. Wicked problems involve intertwined social, economic, and political issues that co-evolve, and stakeholders with heterogeneous views and dynamic reactions in the inter-influence loop.

The COVID-19 pandemic involves not only a biologically novel and evolving virus for which we have no medical cure but also intertwined political, economic, and societal issues, in a global context. Many conflicting and emergent factors are ambiguous and difficult to recognize

E-mail address: luo@sutd.edu.sg.

https://doi.org/10.1016/j.techfore.2021.120602
Received 29 July 2020; Received in revised form 9 January 2021; Accepted 12 January 2021
Available online 19 January 2021

Table 1
Public COVID-19 forecasting initiatives around the world.

Organization	URL	Methods
Imperial College London	https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/	Mechanistic transmission model
University of Geneva, ETH Zürich & EPFL	https://renkulab.shinyapps.io/COVID-19-Epidemic-Forecasting/	Statistical model
Massachusetts Institute of Technology	https://www.covidanalytics.io/projections	Modified SEIR model
Los Alamos National Laboratories	https://covid-19.bsvgateway.org/	Statistical dynamical growth model
The University of Washington, Seattle	https://covid19.healthdata.org/projections	Statistical model
The University of Texas, Austin	https://covid-19.tacc.utexas.edu/projections/	Statistical model
Northeastern University	https://covid19.gleamproject.org/	Spatial epidemic model
University of California, Los Angeles	https://covid19.uclaml.org/	Modified SEIR model
Centers for Disease Control and Prevention, U.S.A.	https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/forecasting.html	Ensemble

and model (Mitroff, 2020; Chen, 2020). Despite the variety of existing forecasting models, they cannot be objectively tested and proved right or wrong due to the lack of a basis on which to define the accuracy of their predictions. In contrast to the "wicked problem", "tame problems" have stable agreements and requirements and can be solved by choosing and applying the correct methods.

With such an understanding of the innate uncertain nature of COVID-19, it would be misleading to focus on accuracy when forecasting its development. However, many ongoing COVID-19 forecasting initiatives still use traditional statistics, e.g., hold-out testing, margins of error, and confidence intervals, to validate or optimize their prediction models. Some ongoing COVID-19 forecasting efforts call confidence intervals "uncertainty bands." Such a calling might be theoretically debatable and cause a *false sense of certainty*. Without a definitive fixed target or ground truth with which to define an "error," talking about errors and confidence intervals makes no sense. Such traditional statistics are only meaningful when we *know* there exists a fixed target but *do not know* to what degree a model hits the target, i.e., a *known unknown*.

The evolution of the current pandemic is affected by many *unknown unknowns* related to the novel and biologically evolving virus, dynamic human behaviors, complex politics, and many other ambiguous, unknown, and emergent factors, as well as their interactions. The uncertainty arising from *unknown unknowns* down the road is essentially *not* quantifiable and is difficult to model for forecasting. A statistical model trained with data from China up to February 2020 could hardly be suitable for predicting later scenarios in the United States, Brazil, India, or other countries with different internal human, social and political dynamics. No COVID-19 forecasting model trained before May 2020 would have incorporated the effects of the Black Lives Matter movement on the pandemic's development, and of other "surprises". No model trained with data before October 2020 would have considered the significantly higher transmissibility of the new coronavirus variants that emerged in the UK and South Africa in September 2020. Given such extreme uncertainty with so many unknown unknowns, it is misleading to validate a prediction model trained on data from a past or different scenario in terms of how accurate it is in a later or different scenario.

3. Our accidental forecasting experience

The challenges arising from the innate uncertainty of the pandemic and its open-ended development do not mean that forecasting based on theory and data is totally useless or that nothing useful can be done. In fact, such uncertainty makes the forecasts of the future even more enticing and meanwhile demand a heuristic and exploratory approach to forecasting. Even if specific forecasts are theoretically always wrong when uncertainty exists, some can be useful in certain ways and to some extent. The concurrent challenges and attractions to forecasting, as well as the heuristics of pandemic forecasting, were evident in our own accidental experiences of data-driven predictions of the development life cycle of COVID-19 in different countries during the period from April 18 to May 11, at the SUTD Data-Driven Innovation Lab (Luo, 2020).

Our predictions were not planned but started purely for our own curiosity on when the pandemic might fade out and were in part driven by our own anxiety from the blindness about the future. On 18 April, we uploaded the initial estimations of the pandemic life cycles of 8 countries (see examples in Fig. 1) on the lab website. Shortly the website received many requests, via the "Contact Us" submission form, from many like-minded people to cover their own countries (beyond the initial eight countries) in our daily updated forecasts. These surprising requests informed us of the shared needs and desires of individuals, companies, public organizations, and governments for the forecasts about the developments of the pandemic in their respective countries, many of which were not covered by the major forecasting initiatives in the United States, Europe, and China (see Table 1).

In response to the growing requests, we added predictions for more countries and continually updated the predictions daily with the latest data. The project escalated accidentally. During this short period of time from April 18 to May 11, the forecasting website received over 7 million unique web visitors from all over the world, with India, Brazil, and United States as the leading source countries. The predictions went viral in social media in India, Brazil, Russia, Indonesia, Singapore, UAE, Turkey, Japan, Germany, among other countries and were widely reported by the media in over 60 countries across the globe. The prediction site received over 37,000 text messages via the "Contact Us" form on the website, the mass majority of which provide positive, encouraging, and constructive comments and feedbacks.

This accidental forecasting experience gave us the grounded learning about the uncertainty and difficulties of COVID-19 forecasting. From the fact that our predictions went viral globally and from our analysis of the media coverage, feedbacks and comments, we developed the empathy toward the people, firms, organizations, and governments around the world about their high anxiety due to the uncertainty and the value of pandemic forecasts to soothe personal mentality and inform organizational planning and government policy. Forecasts are much needed because of the uncertainty but are also challenged by the uncertainty to be done well. This experience together with the feedbacks particularly inspired us about two tactics for pandemic forecasting to address the unknown unknowns associated with COVID-19 and the open-endedness of its future.

4. Forecasting tactics under uncertainty

First, instead of hoping the previous predictions be true or accurate later when the real future comes, we should *monitor these predictions and detect changes in them* over time, using continually updated data. From a traditional accuracy-oriented perspective, a difference between a future prediction and a previous one for the same variable would be considered proof of an error or failure in the prediction model. From a learning-oriented perspective, however, such prediction changes would provide meaningful signals for us to learn about the changes happening in dynamic real-world scenarios, based on the fundamental assumption that the context is uncertain and thus predictions made over time should be different when the real-world scenarios change.

The second forecasting tactic is to *focus on macro patterns and long-term variables*, such as the shape of the total pandemic life cycle (e.g., the curve in Fig. 1) and the total epidemic size or infected population (e.g., the area under the curve) (Debeckera and Modisb, 2020), the inflection date (i.e., the peak of the curve) and the ending of the pandemic

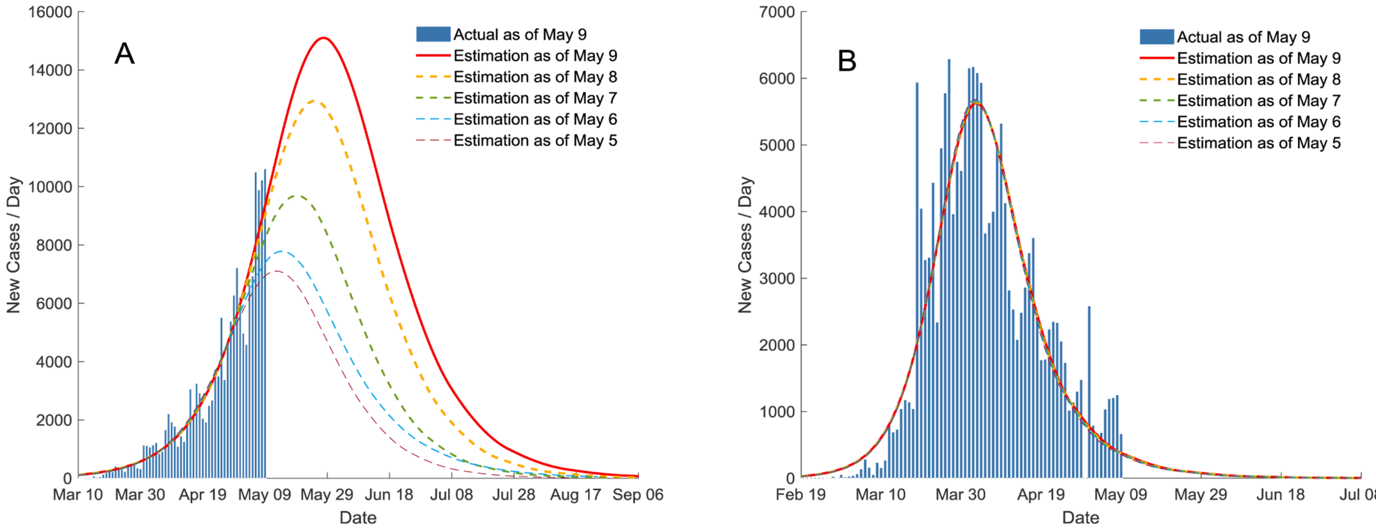

Fig. 1. Predictive monitoring of the pandemic life cycle curves of two countries from May 5 to May 9, 2020. Country names are disguised to avoid political sensitivity. The curves represent the estimated daily new infection cases based on the SIR (susceptible-infected-removed) mechanistic model regressed with actual data reported over 5 consecutive days.[11] The initial segment of the curve is fitted with actual data, shown in solid bars. The remaining segment of the curve is predicted. By plotting the estimated full life cycle curve and actual history together, one can easily sense which phase of the pandemic life cycle a given country is in, when the inflection point (i.e., the peak in the curve) is coming, and when the pandemic might end (i.e., the right tail of the curve). The size of the area below the curve indicates the total expected epidemic size (i.e., the total number of infected and to-be-infected people throughout the entire pandemic life cycle). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(i.e., the right tail of the curve) (Luo, 2020). Such variables correspond to the macro evolutionary characteristic of the pandemic and the total (including past and future) outcome of the past and current scenarios on the ground. This focus on macro patterns and life-cycle-long variables deviates from the desire for accurate predictions of the specific numbers of infection cases in the next day, week, or month in the short term, and may provide forward-looking signals on the possible future scenarios and total outcomes as a result of the actual present scenario and its change.

For example, in Case A in Fig. 1, the uplifting of the continuously predicted pandemic life cycle curves over 5 consecutive days may suggest that the situation on the ground was worsening and resulting in more future infections down the road, later peak dates, and a longer pandemic life cycle. The detection of such changes in predictions via the monitoring of predictions provides warning signals to initiate responsive actions. However, from an accuracy or validation perspective, the volatile changes which we discover here as meaningful future-informed signals would be viewed as negative prediction errors.

In contrast, the serial predictions in Case B were hardly changing and indicate the situation on the ground was stable. However, the stability might be still undesirable, if the stably estimated macro patterns reveal a long tail or total life cycle and a large, expected population to be infected. Such discoveries may inform the government to proactively implement more stringent controls to improve the ground situations and result in a better future course of development. If the responsive or proactive actions are effective, the later updated predictions of the pandemic curves would show bending and shortening of the estimated overall life cycle curves from previously estimations. Therefore, monitoring such real-time dynamic signals on the possible future scenarios and estimated macro consequences, in addition to monitoring the actual history and data to date, may initiate, guide, and assess responsive and proactive actions.

5. Predictive monitoring

Therefore, we propose to monitor the changes in the predictions, which are continually updated with the latest data daily, and focus on the predictions of macro patterns and long-term variables regarding the total pandemic life cycle. Taken together, we propose and advocate the "*predictive monitoring*" paradigm, which synthesizes predictions and monitoring, for heuristic learning of the uncertain and open-ended developments of the pandemic. *Predictive monitoring* differs fundamentally from the traditionally defined prediction and monitoring paradigms. Table 2 presents a taxonomy.

Traditional prediction practices attempt to make a prediction now that can be accurate about the future (assuming the future is fixed) and are made like weather forecasts, wherein future weather cannot be changed by human stakeholders. Such practices are more meaningful when uncertainty is low in the context. Traditional monitoring practices track the actual history to date and can inform and guide reactive and responsive actions in an uncertain context. But they do not inform us about the future consequences of what has been taking place. To explore possible future scenarios in a highly uncertain context, the practice of *predictive monitoring* would track predictions of the same long-term variables made over time to detect and generate forward-looking signals on the ambiguous changes in present real-world scenarios together with their possible future consequences.

Therefore, *predictive monitoring* complements traditional prediction and traditional monitoring practices. For countries that are in early stages of the pandemic life cycle (e.g., Case A in Fig. 1), predictions about the remainder of the total pandemic life cycle curve, total accumulative cases and ending dates will be more tempting. However, specific and individual predictions will be inherently less relevant to the "real future" to come, because the actual data cover only a smaller and early portion of the total life cycle, real-world scenarios will evolve and change, and surprises will emerge. Predictive monitoring is especially valuable for such cases. In contrast, for countries approaching the

Table 2
A taxonomy of predictive monitoring, prediction, and monitoring.

		The Value It Delivers Future-Informed	Past-Informed
The Context It Suits	Wicked Problem	*Predictive Monitoring*	*Monitoring*
	Tame Problem	*Prediction*	

ending phase with effective controls of the pandemic (e.g., Case B in Fig. 1), specific predictions will be more trustworthy but less desired. In such cases, the data-trained model and estimated life cycle curve are more about explaining the history, whereas monitoring of daily actual cases remains crucial as uncertainty still exists.

However, a new epidemic wave might emerge if governments and individuals lift controls and relax disciplines dramatically, especially when the pandemic is still prevalent in other countries and definitive medical cures do not exist. This is the case for many European and Asian countries, which entered a scenario depicted by Case B in Fig. 1 in summer 2020 and thereafter eased controls and social distancing and allowed a new epidemic wave to break out in fall 2020. For countries in the growth phase of the new epidemic wave (e.g., a scenario depicted by Case A in Fig. 1), predictive monitoring becomes valuable again.

6. Summary

In sum, the ongoing COVID-19 pandemic is a "wicked problem," and its innate uncertainty challenges accuracy-oriented forecasting efforts. The viruses constantly evolve and change through mutations and multiple COVID-19 variants have been emerging in the UK, South Africa, and other regions. For many countries, human behaviors, government policies, and other social, economic, and political factors are still dynamic and make predicting the course of the pandemic challenging. To address uncertainty proactively and make sense of pandemic forecasting from a heuristic learning-oriented perspective, we propose and advocate the predictive monitoring paradigm. Predictive monitoring aims to capture changes in continually updated predictions of the same macro patterns and long-term variables and infer such dynamic forward-looking signals back to the changes and situations on the current ground and to required interventions. It contrasts with the traditional prediction that aims for accuracy and assume a fixed future and complements the traditional monitoring that tracks the actual history only. Predictive monitoring harnesses the spirits and methods of both prediction and monitoring to inform heuristic decisions and future-minded planning under extreme uncertainty. Its applicability goes beyond the COVID-19 pandemic.

References

Box, G.E.P., 1976. Science and statistics. J. Am. Stat. Assoc. 71 (356), 791–799.
Chen, Z., 2020. COVID-19: a revelation - A reply to Ian Mitroff. Technol. Forecast. Soc. Change 156, 120072. https://doi.org/10.1016/j.techfore.2020.120072.
Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Piontti, A.P., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M.E., Longini Jr., I. M., Vespignani, A., 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 368 (6489), 395–400.
Debeckera, A., Modisb, T., 2020. Poorly known aspects of flattening the curve of COVID-19. Technol. Forecast. Soc. Change. https://doi.org/10.1016/j.techfore.2020.120432 in press.
Ferguson, N.M., Imperial College COVID-19 Response Team, 2020. Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand. Imperial College London. March 16. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf.
Grey, S., MacAskill, A., 2020. Special report: johnson listened to his scientists about coronavirus - but they were slow to sound the alarm. Reuters. April 7. https://reut.rs/2Rk3sa7.
IHME COVID-19 Health Service Utilization Forecasting Team, 2020a. Forecasting COVID-19 Impact on Hospital Bed-Days, ICU-Days, Ventilator-Days and Deaths by US State in the Next 4 Months. MedRxiv. https://doi.org/10.1101/2020.03.27.20043752. March 30.
IHME COVID-19 Health Service Utilization Forecasting Team, 2020b. Forecasting the Impact of the First Wave of the COVID-19 Pandemic on Hospital Demand and Deaths for the USA and European Economic Area Countries. MedRxiv. https://doi.org/10.1101/2020.04.21.20074732. April 26.
Jewell, N.P., Lewnard, J.A., Jewell, B.L., 2020. Caution warranted: using the institute for health metrics and evaluation model for predicting the course of the COVID-19 pandemic. Ann. Intern. Med. https://doi.org/10.7326/M20-1565. April 14.
Kissler, S.M., Tedijanto, C., Goldstein, E., Grad, Y.H., Lipsitch, M., 2020. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 368 (6493), 860–868.
Luo, J., 2020. When Will COVID-19 End: Data-Driven Predictions. Data-Driven Innovation Lab. Singapore University of Technology & Design. April 30. https://www.persi.or.id/images/2020/data/covid19_prediction_paper.pdf.
Marchant, R., Samia, N.I., Rosen, O., Tanner, M.A., Cripps, S., 2020. Learning As We go: An examination of the Statistical Accuracy of Covid19 Daily Death Count Predictions. MedRxiv. https://doi.org/10.1101/2020.04.11.20062257. April 17.
Mitroff, I.I., 2020. Corona virus: a prime example of a wicked mess. Technol. Forecast. Soc. Change 157, 120071. https://doi.org/10.1016/j.techfore.2020.120071.
Petropoulos, F., Makridakis, S., 2020. Forecasting the novel coronavirus COVID-19. PLoS ONE 15 (3), e0231236.
Ray, E.L., Wattanachit, N., Niemi, J., Kanji, A.H., House, K., Cramer, E.Y., Bracher, J., Zheng, A., Yamana, T.K., Xiong, X., Woody, S., Wang, Y., Wang, L., Walraven, R.L., Tomar, V., Sherratt, K., Sheldon, D., Reiner Jr, R.C., Prakash, B.A., Osthus, D., Li, L. M., Lee, E.C., Koyluoglu, U., Keskinocak, P., Gu, Y., Gu, Q., George, G.E., España, G., Corsetti, S., Chhatwal, J., Cavany, S., Biegel, H., Ben-Nun, M., Walker, J., Slayton, R., Lopez, V., Biggerstaff, M., Johansson, M.A., Reich, N.G., 2020. Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the U.S. MedRxiv. https://doi.org/10.1101/2020.08.19.20177493. August 22.
Resnick, B., 2020. The White House Projects 100,000 to 200,000 Covid-19 Deaths. Vox, March 31. https://www.vox.com/science-and-health/2020/3/31/21202188/us-deaths-coronavirus-trump-white-house-presser-modeling-100000.
Rittel, H.W.J., Webber, M.M., 1973. Dilemmas in a general theory of planning. Policy Sci. 4 (2), 155–169.
Woody, S., Tec, M.G., Dahan, M., Gaither, K., Lachmann, M., Fox, S., Meyers, L.A., Scott, J.G., 2020. Projections for First-Wave COVID-19 Deaths Across the US Using Social-Distancing Measures Derived from Mobile Phones. MedRxiv. https://doi.org/10.1101/2020.04.16.20068163. April 26.
Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., He, J., 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J. Thorac. Dis. 12 (3), 165–174.

Jianxi Luo is a tenured Associate Professor with SUTD, Director of Data-Driven Innovation Lab, and Director of SUTD Technology Entrepreneurship Programme. Prof. Luo holds a PhD in Engineering Systems (Technology Management and Policy track) and S.M. in Technology Policy from MIT. He was a faculty member at New York University and general chair of INFORMS Technology Innovation Management & Entrepreneurship Section. He is on the editorial boards of IEEE Transactions on Engineering Management, Research in Engineering Design, and Design Science (Associate Editor). His-research focuses on data-driven innovation methods, tools, and systems, and has published over 120 academic articles.

[1] Public data were from Our World in Data (https://github.com/owid/covid-19-data/tree/master/public/data) and open source regression codes were from Milan Batista (https://www.mathworks.com/matlabcentral/fileexchange/74658-fitviruscovid19).