

Editorial: Epidemics and forecasting with focus on COVID-19

1 Forecasting and decision-making for COVID-19

Despite a few signs late December 2019, when the first cases of COVID-19 were officially reported in China, and early January 2020 when first cases of COVID-19 were reported outside of China (indicating the further spreading of the infectious disease), we were nearly all caught by surprise as these early events developed into a lasting worldwide pandemic. Rapidly, different institutions and websites starting to collect and communicate data related to the pandemic, mainly related to the number of cases, the number of tests being performed, the basic reproduction number R_0 , the number of hospitalizations, etc. And rapidly too, many looked into forecasting of some of these variables, especially the number of cases, hospitalizations and use of intensive care unit (ICU) beds. This was due to the fact that many decision-making problems required such forecasts, in relation to potential confinements and the use of varied restrictions, as well as health and safety measures. Over the following period, and still today 2 years after the start of the pandemic, COVID-19-related forecasts are at the core of governmental actions and widely communicated to a broad audience.

Since aiming to be a leading journal focusing on the science and practice of forecasting, the *International Journal of Forecasting*, and more broadly the International Institute of Forecasters (IIF), looked into ways to foster the development of relevant knowledge for forecasting in relation to COVID-19. Obviously, substantial expertise already existed out there thanks to the decades of work of epidemiologists in relation to forecasting. The mathematical modelling of infectious disease is an old scientific field, arguably traced back to the works of Bernoulli in 1760, while the compartmental models (e.g., the Susceptible, Infectious, or Recovered – SIR model, and the Susceptible, Exposed, Infectious, or Recovered – SEIR model, as well as their many variations) that have been widely used over the last 2 years were first proposed in the 1920s. This existing expertise was invaluable for the development of relevant models that were used for the forecasting of the development of COVID-19 worldwide since early 2020. However, as for many other application areas, the science and practice of forecasting are evolving tremendously thanks new possibilities offered by the large amounts of data being collected, advances in mathematical modelling, statistics and machine learning, as well as computing power. Forecasting has also developed into a democratic science, with strong interaction between scientists and non-scientists, between forecasters and forecast users. In a way similar to open forecast competitions where expert and non-experts do their best to produce forecasts for set forecasting problems and relevant benchmarks, many scientists and non-scientists engaged in developing and applying forecasting approaches for COVID-19, which were evaluated in real-time and which evolved over the last few years. All in all, this availability of data and the interest of many to contribute to the development of forecasting approaches for COVID-19 led to a buzzing environment with a wealth of original ideas, proposals, benchmarking exercises and heated discussions (for instance on social media).

In collaboration with the IIF, we hosted a number of forecasting-related blog posts at forecasters.org. We additionally worked towards the assembling of relevant knowledge in the form of a special section to be published in the journal, and which is gathered here.

2 Contents and main messages

As an opening to this special section, Pinson and Makridakis introduce the scientific debate that was organized between Nassim N. Taleb (co-authors: Yaneer Bar-Yam and Pasquale Cirillo) and John P. Ioannidis (co-authors: Sally Cripps and Martin A. Tanner). At the time when diverging voices were heard about how to rely on science and facts to support decision-making at the early stage of the pandemic development, it was important to give the time for scientists and thought leaders to debate about, and argue for, the role forecasting should (or should not) play in fighting the pandemic. The approach for such a respectful and insightful scientific debate is, in our opinion, a good way to move forward when the usual time and room for scientific and academic thinking is being upset by the rapid unfolding of real-world events.

Taleb, Bar-Yam and Cirillo ask themselves an important question about our ability to predict the development of a pandemic like that induced by COVID-19, and especially if the aim is to present future estimates in the form of single-point forecasts. In a general manner, it is broadly accepted today that forecasts should be thought of and communicated within a probabilistic framework. The case is even stronger for processes and variables that are fat-tailed, with highly asymmetric and uncertain loss functions. The paper also aims at drawing more general conclusions about the interplay between forecasting and risk management, as well as about forecasting while we concurrently act upon those processes we aim at predicting.

In parallel, Ioannidis, Cripps and Tanner offer us important insights on epidemic forecasting, and more specifically for the case of COVID-19. Obviously, epidemic forecasting is a tremendous challenge with high stakes, and obviously forecasts are wrong. It is therefore of utmost importance to look back, analyse what happened, see what can be learned, and look for ways to move forward in the right direction. The paper does that with a lot of clarity and ingenuity. In addition, in the appendix of the paper, Ioannidis gives us what he refers to as a “fool’s confession” – he should be commended for his honesty and his ability to reflect on his own views.

Petropoulos, Makridakis, Stylianou share with us their experience in developing and using operationally an approach to the short-term development of the pandemic, in terms of the number of confirmed cases and deaths, from (almost) its very start. Their argument is that if focusing on the short-term only (i.e., 10 days ahead), while issuing and communicating forecasts in a probabilistic format (as predictive densities), it may indeed be possible to readily employ ideas and tools from time-series analysis to do so. Specifically here, they used the non-seasonal multiplicative error and multiplicative trend exponential smoothing model. They provide us with a rigorous and thorough analysis of the quality of the forecasts, also expressing thoughts about limitations, lessons learnt and paths for improvement.

Doornik, Castle and Hendry describe a method that relies on similar premises and strategy, i.e., that the recent information in collected data for confirmed cases and deaths may be informative to issue short-term projections based on statistical modelling tools only. There, they use Local Averaged Time Trend Estimation (abbreviated LATTE), with a few adaptations to the specifics of the application to COVID-19 data. The approach allows to issue forecasts as conditional expectations, but also to derive prediction intervals for a given nominal coverage rate (they consider 80% in their operational application and empirical investigation). They also provide extensive application results and discussion of their approach compared to the use of SIR models. An interesting angle of the paper, relating to the practice of forecasting, is that the authors also clearly explain how they adapted their approach and strategy as the pandemic developed, and why.

While the previous 2 papers concentrated on using local information only, the proposal of Medeiros and co-authors is rooted in the idea such that an infectious disease like COVID-19 developed over various areas of the worlds with heterogeneous delays. Therefore, data from countries that witnessed earlier substantial development of the pandemic could be useful to issue better forecasts for other regions or countries that experienced a delay. Consequently, they use an error correction model and a LASSO estimation approach as a way to select the relevant information among all potential data origins (regions, countries, time lags) considered. They additionally employ a sliding-window re-estimation approach to accommodate the non-stationarity in the underlying stochastic

processes. They finally provide extensive applications results, and related analysis, for 4 countries out of the 45 they initially studied: Brazil, Chile, Mexico, and Portugal.

Achterberg and co-authors somewhat push that kind of approach even further by looking at the propagation of infectious diseases and the dynamics of COVID-19 on a network, hence directly accounting for the connection among regions and countries, and how they may influence each other. They eventually describe the so-called Network Inference-based Prediction Algorithm (NIPA), which relies on a combination of machine learning components and a SIR-like model, also possibly accounting for linkage and interaction among different regions and countries. Comparisons of various approaches are performed for different case-studies (regions and lead times), allowing to discuss the interest of the various components of the NIPA approach, as well as the overall quality of the forecasts obtained.

As an alternative, Chiang, Liu and Mohler bring forward the idea of using Hawkes processes (for event clustering) and spatial covariates, which they argue can be seen as a stochastic generalization of the compartmental models (e.g., the SIR and SEIR ones). They introduce an Expectation-Maximization (EM) algorithm for estimation in such Hawkes process models, while discussing its connection to a dynamic approach to the estimation of the basic reproduction number R_0 . Their approach additionally allows to accommodate a number of relevant additional covariates, for instance related to demography, health, habits and mobility patterns at county level in the US. They use a case-study application to discuss the relative importance of such covariates.

Finally, as a closing article, Coughlan de Perez and co-authors draw a parallel between forecasting and decision-making under uncertainty related to weather and climate, with the situation we have been through with the COVID-19 pandemic and its development. Their perspective paper identifies 3 major areas of focus, which relates to the fact that forecasts are there to be used for decision support, to the importance of conveying uncertainties, and to the understanding of vulnerabilities. Regarding that final point, they discuss the fact that the drift of weather forecasting from the prediction of certain meteorological variables (e.g., wind and temperature) towards impact-based forecasting (e.g., storm and flood warnings) may be seen as example of how forecasting is to evolve by further integrating the appraisal of the variety of potential loss functions of a broad range of stakeholders.

By assembling this special section, we hope that the International Journal of Forecasting will make a substantial contribution to the way we see epidemic forecasting, and to the way we move forward after the recent developments in the COVID-19 pandemic.

Acknowledgements

Acknowledgements are due to all who contributed to this special section, as well as the many discussions prior and after its preparation. I am grateful to all Editors of the International Journal of Forecasting for triggering this initiative and for agreeing on its format. Additional thanks go to Pam Stroud at the International Institute of Forecasters (IIF) for helping with editing, formatting and posting some of the blog posts related to that initiative (at forecasters.org), as well as to Spyros Makridakis for helping with the scientific debate between Nassim N. Taleb (Yaneer Bar-Yam and Pasquale Cirillo) and John P. Ioannidis (Sally Cripps and Martin A. Tanner).

Pierre Pinson
Editor-in-Chief
International Journal of Forecasting