

Journal Pre-proof

COVID-19: Forecasting confirmed cases and deaths with a simple time-series model

Fotios Petropoulos, Spyros Makridakis, Neophytos Stylianou

PII: S0169-2070(20)30186-2

DOI: <https://doi.org/10.1016/j.ijforecast.2020.11.010>

Reference: INTFOR 4864

To appear in: *International Journal of Forecasting*



Please cite this article as: F. Petropoulos, S. Makridakis and N. Stylianou, COVID-19: Forecasting confirmed cases and deaths with a simple time-series model. *International Journal of Forecasting* (2020), doi: <https://doi.org/10.1016/j.ijforecast.2020.11.010>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

COVID-19: Forecasting confirmed cases and deaths with a simple time-series model

Fotios Petropoulos¹, Spyros Makridakis², and Neophytos Stylianou^{3,1}

¹ School of Management, University of Bath, UK

² Institute for the Future (IFF), University of Nicosia, Nicosia, Cyprus

³ International Institute for Compassionate Care, Cyprus

Time is running out to prepare for the next pandemic.

We must act now with decisiveness and purpose.

Michael Osterholm, 2005

Abstract

Forecasting the outcome of outbreaks as early and as accurately as possible is crucial for decision making and policy implementations. A significant challenge faced by forecasters is that not all outbreaks and epidemics turn into pandemics making the prediction of their severity difficult. At the same time, the decisions made to enforce lockdowns and other mitigating interventions versus their socioeconomic consequences are not only hard to make, but also highly uncertain. The majority of modeling approaches to outbreaks, epidemics, and pandemics take an epidemiological approach that considers biological and disease processes. In this paper, we accept the limitations of forecasting to predict the long-term trajectory of an outbreak, and instead, we propose a statistical, time-series approach to modelling and predicting the short-term behaviour of COVID-19. Our model assumes a multiplicative trend, aiming to capture the continuation of the two variables we predict (global confirmed cases and deaths) as well as their uncertainty. We present the timeline of producing and evaluating 10-day-ahead forecasts over a period of four months. Our simple model offers competitive forecast accuracy and estimates of uncertainty that are useful and practically relevant.

Keywords: COVID-19; decision making; exponential smoothing; pandemic; time-series forecasting; uncertainty.

1. Introduction

The impact of COVID-19 and the measures needed to fight the pandemic present a gruelling challenge to policymakers (Ferguson et al., 2020) with profound implications for humanity. The choices are grim. Governments are forced to choose between (i) lockdowns to stop the spread of the virus and reduce the number of deaths while tolerating the grave economic consequences that may lead to economic depression, and (ii) allowing business as usual to keep the economy going while enduring loss of life and human suffering that many find unacceptable. By 31 May 2020, the number of reported, global confirmed COVID-19 cases surpassed the 6 million mark while the number of deaths attributed to the pandemic had reached 369,000. We also observe huge differences among countries, with over 820 deaths per million inhabitants in Belgium but single-digit lows in many others (Statista, 2020), and even more significant variations in the per age mortality rates.

Setting aside the reliability of these numbers, national policies concerning lockdowns and social distancing and how soon they are introduced have been a critical factor in both infections and death rates. A study estimated the number of deaths by 3rd May would have been lower by 36,000 if the interventions to control the virus in the USA had started a week earlier, and by 54,000 if the interventions had started two weeks before (Pei et al., 2020). The type and timing of interventions have been a decisive factor in the spread of the disease and the subsequent death rates. However, there is no agreement about their exact influence and what could have been done to minimise human suffering and loss of lives while avoiding the disastrous economic implications of lockdowns.

In this introduction we explore four issues that could have reduced the negative impact of the decisions made. First, could the COVID-19 pandemic have been predicted? Alternatively, could the world have been better prepared to face it even though its exact timing and impact was not known? Second, could the measures taken have been more effective if lessons learned from previous pandemics had been applied to this one? Third, what additional information could have been provided to policy and decision makers to improve their decisions? Finally, even if "perfect" information had been available, could ideological beliefs have been put aside to decide which measures to adopt in a more rational way?

1.1 Predicting and being prepared for pandemics

Cirillo and Taleb (2020) analysed 72 major pandemics spanning over 2,500 years and showed that they are fat-tailed in terms of duration and fatalities (see their Table 1), posing significant risks for humanity. Therefore, forecasting the specific impact of a pandemic beforehand would have been practically impossible and too uncertain. Some last only months, killing less than a thousand people, while others can endure for more than 20 years, annihilating more than a third of the world's population.

The latest pandemic, the Spanish flu, started in 1918 and killed an estimated 58.5 million people (194 million in terms of the size of the population in 2020). It was followed by the HIV/AIDS pandemic that is still active in some form, as well as the Asian flu and the Hong Kong flu that killed 3, 2 and 1 million people respectively (62, 5 and 2 million respectively in terms of the population size in 2020). In comparison, COVID-19 has resulted in 1.2 million deaths until now and has caused a great amount of human pain and economic destruction. All pandemics (no matter how small) can produce great upheaval by disturbing our everyday lives and affecting our economic activity. Worse, we cannot predict when they will hit, how long they will last or their human and economic impact, generating huge uncertainty and psychological and economic upheaval for a good number of people (Lilla, 2020).

There are strong suggestions by epidemiologists and other experts that pandemics are inevitable (Carras, 2020; Osterholm, 2005; Taleb, 2008) and that there is an urgent need to be prepared to face them, whenever time they might appear (Osterholm & Olshaker, 2020). There is also the assertion that COVID-19 is not "the big one" and that humanity must be prepared to face a much worse pandemic, most likely a novel influenza virus with the same devastating impact as the Spanish flu pandemic of 1918, which circled the globe two and a half times over more than a year, in three recurring waves, killing many more people than the brutal and bloody war that preceded it (Osterholm & Olshaker, 2020).

Pandemics are not the only natural events that cause considerable human and economic suffering. Earthquakes, hurricanes, floods, avalanches and tsunamis are of a similar kind, leaving traces of destruction in their path. In contrast to pandemics and tsunamis, the remaining natural disasters are more frequent, which results in more experience gained in order to be prepared to deal with them as efficiently as possible and minimise their damages.

There are different types of preparation for natural disasters whose occurrence cannot be predicted, and the example of Japan's history with earthquakes is worth imitating. Japan accounts for around 20% of earthquakes around the world with a magnitude 6.0 or more, where around 1,500 earthquakes strike this island nation every year and minor tremors occur nearly every day. Japan is prepared to face these earthquakes by having implemented a combination of long and short-term measures to be able to function routinely and minimise loss of life and economic hardship. The long-term measures include earthquake-resistant buildings/structures and infrastructure projects to avoid floods and transportation problems (for instance, its famous high-speed bullet trains stop automatically at the first sign of an earthquake). There is a constant effort to raise awareness and educate people, including young children, on what to do and how to face an earthquake. Although the analogy between earthquakes and pandemics may not seem straightforward, there is a lot to be done to be better prepared by educating people on the value of masks and the need for social distancing, as well as having plans in place to restrict travel and protect the elderly and those in nursing homes.

1.2 Learning from past pandemics

Dangerous epidemics are on the rise around the world. Their numbers have increased nearly fourfold over the past 60 years, while their yearly number has more than tripled since 1980 (Walsh, 2017). Pandemics are less frequent than epidemics, but their impact is much more severe. Epidemiologists contribute their rise to the increased globalisation and the more significant social interaction among people around the world and suggest greater efforts must be made to be better prepared to face them, even though they always come as a surprise.

During the 20th century, there were four pandemics and each recorded more than one million deaths. From those, the Spanish flu (Taubenberger & Morens, 2006) killed an estimated 58.5 million while HIV/AIDS, which still exists today, has caused more than 30 million deaths. The Spanish flu came in multiple, distinct waves over a period of two years with considerable differences in death rates between Europe and USA and between various US cities, and like COVID-19, these rates depended on how early during the pandemic interventions were made to reduce its spread (Bootsma & Ferguson, 2007). The pattern of HIV/AIDS has been quite different from that of the Spanish flu; while it has lasted for over 50 years, the number of infected with HIV/AIDS peaked at the end of 1999, declining since then, and the number of deaths has been declining since 2006 as more information about the disease and more effective medical treatments have become available.

So far in the 21st century, no pandemic has killed more than one million people, although COVID-19 has just now passed this mark. The infrequency of pandemics has contributed to underestimating its severity and economic impact as lockdowns and social distancing prevented a good part of normal economic activity, increasing unemployment, and probably leading to a major recession that will be influenced by the severity of a second, or possibly third wave of the virus hitting the world, thus requiring new lockdowns and social distancing measures to be imposed. In South Korea, for instance, schools were obliged to close again after the largest spike in weeks (Mahbubani, 2020).

If we accept that pandemics will hit us again in the future without warning, we must be prepared to face them as effectively as possible, for instance, by banning travelling early, immediately making facemasks obligatory, and imposing social distancing among other measures.

1.3 Additional information that would have improved decisions

The infrequency of pandemics contributes to human bias, making us believe the illusion (Langer, 1975) that they will not affect us and there is no need to take the difficult actions needed to minimise their impact. As mentioned earlier, if the lockdown in the USA had been initiated a week earlier, there would be 36,000 fewer deaths. At the same time, it would have been a week of reduced economic activity with many people unable to work.

Ioannidis (2020), in a paper titled "A fiasco in the making? As the coronavirus pandemic takes hold, we are making decisions without reliable data" written early during the pandemic, argues that more reliable information is needed to guide decisions and actions of monumental significance and to monitor their impact. He continues saying that given the uncertainty involved, a reasonable estimate for the case fatality ratio in the general U.S. population could vary from 0.05% to 1% and that at the lower range, it would be the same as the seasonal flu. Moreover, Ioannidis raises the issue of the great difference in death rates according to age, particularly for people over 75 and for those with existing health conditions, and whether the average rate should be adjusted to reflect those differences in policy decisions. He does not question imposing lockdowns and social distancing, but rather suggests that better data is needed to be able to make these trillion dollar decisions. For instance, he suggests spending some money on sampling the general population instead of not tracking the fatality rates or other vital statistics. Others disagree with Ioannidis (Cirillo & Taleb, 2020), claiming that the risks are highly asymmetric, that no precise information is needed to decide that lockdowns and social distancing are necessary, and if not implemented will lead to great suffering and the loss of innumerable lives that will cost a great many times as much to correct in the future. It is the precise aim of the debate between Taleb and Ioannidis published in this special issue to answer the question of how to more effectively deal with pandemics.

1.4 Ideological preferences

What is the cost of human life (Rogers, 2020)? Can such a cost be compared to the economic damages caused by lockdowns and social distancing? New York Governor Andrew Cuomo in a briefing stated (Eyewitness News, 2020): "To me, I say the cost of a human life, a human life is priceless. Period." Others, however, worry about the impact of lockdowns and social distancing on the economy and are citing facts such as the worst drop in the US Gross Domestic Product since 2008 and the 33.5 million people who have filed for unemployment since March when the lockdowns took effect. They also ask, what will happen if the economy does not open up soon and how to deal with the bankruptcies that will become inevitable and the many people without any income to pay their rent and support their families?

In addition, they question what will happen if there is a second and possibly a third wave of the virus? Could a new lockdown be imposed? Clearly, these are not easy questions to answer, as Rogers (2020) concludes in his article "How Much Is a Human Life Actually Worth?" that attempts to estimate a number to this question, without much success. He suggests that instead of putting a number on human life, "what researchers would like to know is which specific interventions are most successful stopping the virus and have the least impact on people's economic lives." Better management of the pandemic could be aligned with economic interests; enhancing a culture of duty and responsibility towards protecting each other (through wearing masks for example) can lead to lower transmission rates and delayed, if not avoided, lockdowns. However, there is also the issue of the unequal burden of the pandemic affecting the poor much more than the rich (Dahir, 2020).

Such uncertainty/disagreement brings us back to square one, where opinion, heavily coloured by ideological preferences, influences decisions. Moreover, political considerations are heavily weighted by politicians who know that voters are influenced by the state of the economy and unemployment, and will make decisions and take actions to improve the economy and therefore, their chance of being reelected. Unfortunately, the situation cannot substantially change from having to decide between two grim choices until a successful vaccine is found, and that is not likely to happen until more than a year from now. Until then, the debate between Taleb and Ioannidis, published in this special issue of IJF, covers these topics and provides the perspective of the two top researchers and their teams.

Although medium and long-term forecasting does not seem to be of much value for policy and decision makers, short-term forecasting can be useful and will be implemented in the second part of this paper, after a literature review of past pandemics as well as the present one is provided, the mathematical and judgmental models used to forecast them are discussed, and their usefulness and accuracy, as well as their advantages and limitations, are presented.

1.5 Our contribution

The major purpose of this paper is to propose a short-term forecasting model to predict the confirmed cases and deaths of the coronavirus over a period of four months and to evaluate its accuracy and usefulness. This is done by forecasting both variables for 10 days ahead and repeating the process 12 times in a unique, live forecasting experiment with vast potential implications for planning and decision making, by providing objective forecasts for the confirmed cases and deaths whose accuracy can be traced and compared over a period spanning 120 days. We extend the study by Petropoulos and Makridakis (2020) to more rounds of forecasts and two variables, while analysing the forecast accuracy in terms of different planning horizons and the performance of the prediction intervals. We compare the performance of our approach to other publicly available forecasts. Finally, we provide country-specific forecasting for three closely related countries that implemented different levels of interventions (Denmark, Norway, and Sweden).

The remainder of the paper consists of four parts. The first part of the paper provides a review of methods and models used to forecast the current pandemic. The second part describes the proposed short-term forecasting model that uses 12 rounds of predicting confirmed rates and deaths for the entire world and some select countries, as well as their accuracy and value to policy and decision makers. The third part is a discussion of our ability to forecast pandemics and the uncertainty associated in doing so, what we can learn from them and how we can improve our ability to deal with future ones. The final part provides concluding remarks.

2. Literature review

As unknown infectious diseases emerge, creating an outbreak which leads to an epidemic and ultimately a pandemic, researchers try to use modelling exercises to describe observed patterns and try to predict or forecast those patterns in the future so that public health services can be prepared and plan their responses. Epidemic forecasting, specifically, is of paramount importance to epidemiologists, healthcare providers and health policy makers.

With the advancement of data science, numerous forecasting methods have been proposed in the field of epidemic forecasting. Most forecasting modelling methods employ similar approaches. Most of the models are mathematical, using time series epidemiological data in combination with some additional parameters in order to make informed predictions.

There are various predominant approaches to forecasting diseases; namely statistical, mechanistic and judgmental methods.

Statistical methods model disease outbreaks by identifying time series patterns in historical data, but do not account for disease transmission dynamics directly and do not take into account any biological processes (Brooks et al., 2015; Kandula et al., 2018; Wang et al., 2015). The time series models try to predict epidemiological behaviours by modelling historical surveillance data. Many researchers have applied different time series models to forecast epidemic incidence in previous studies. Some of these models use exponential smoothing (Tseng & Shih, 2019), generalised regression (Imai et al., 2015) multilevel time series models (Spaeder & Fackler, 2012), and autoregressive integrated moving average (ARIMA) models (Li et al., 2012). Based on their methodological simplicity and lack of complex mathematics, simple time series models are considered relatively easy methods to explain to end users which improves the trust in the model's outcomes and thus its usage.

Epidemiological methods try to model disease states taking into account the biological and disease processes such as disease transmission processes and individual and population variables (Hyder et al., 2013; Shaman et al., 2013; Shaman & Karspeck, 2012), thus making them inevitably more complex and more computationally challenging. The most common model in epidemiological forecasting of infectious diseases is the SEIR (Susceptible - Exposed - Infected - Recovered) model and its variation the SEIRD model (Piccolomini & Zama, 2020), which has a deaths compartment added as well. The population is divided into the appropriate compartment, and they move between compartments during different stages of the disease. This modelling approach takes into account the infection and recovery rate in the population (Becker & Grenfell, 2017; Kermack et al., 1927). A different group of methodologies used in disease forecasting are machine learning models utilising artificial neural networks and support vector algorithms, which are gaining ground lately in forecasting infectious disease incidence (Philemon et al., 2019; Tapak et al., 2019).

Another noticeable forecasting approach is the use of judgment. Collective human judgment is known to have good predictive power but is also influenced by human biases. Research showed that in some cases, it can match or even exceed most statistical and epidemiological methods (Farrow et al., 2017). Over the years, the quality of judgmental forecasts has improved by implementing well-structured and systematic approaches (Good Judgment, 2020). Judgmental forecasting is the only forecasting option, due to the complete lack of historical data, in cases such as during completely new and unique market conditions or when an outbreak occurs from a previously unknown novel pathogen. In these situations, judgment may be applied in order to forecast the effect of such a pathogen or any policies that will be decided in the absence of historical precedents. Judgmental forecasting is also useful in situations where data is not collected in a timely manner or when the data is incomplete. For these reasons, judgmental forecasting may be applicable to a good degree to the COVID-19 pandemic.

Forecasting models can also be combined into an ensemble. This combination can be many models of the same methodological approach, for example, many ARIMA models, or a combination of mixed methodological approaches. An ensemble that combines different forecast methods, including both statistical and epidemiological models, could improve accuracy while reducing forecast uncertainty, outperforming single methods. Ensemble weather predictions, as well as ensemble approaches to infectious disease forecasting, have reported promising improvements (Krishnamurti et al., 1999; Smith et al., 2017; Viboud et al., 2018).

Regardless of the choice of the forecasting method employed, every study should report on the model's performance (Tabataba et al., 2017) which should include the accuracy of the point forecasts but also the performance of the prediction intervals or quantile forecasts (i.e., the model's ability to capture the uncertainty around the point forecasts).

Before COVID-19, the 2014-2015 West African Ebola epidemic was one of the most heavily modelled outbreaks in history (Chretien et al., 2015). A growing number of models have been developed by health care systems, academic institutions, consulting firms and others, to help forecast COVID-19 cases and deaths, medical supply needs, including ventilators, hospital beds and intensive care unit (ICU) beds, timing of patient surges and more. For the purpose of this paper we will focus on methods and techniques used for the forecasting of the current COVID-19 pandemic, separated into models that focus on national and international outcomes.

2.1 Forecasting models: Country-specific

Since the beginning of the pandemic, many forecasting models have been developed with global, national, or even regional forecasting outputs. Sujath et al. (2020) proposed a model that could be used to forecast the spread of COVID-19 in India using linear regression (LR), Multilayer perceptron (MLP) and Vector autoregression (VAR). They concluded that MLP provides better results compared to LR and VAR, but no accuracy or uncertainty metrics were provided. Gupta and Pal (2020) have used ARIMA and exponential

smoothing techniques to forecast infected cases and deaths in India and its states, as well as the South Asian Association for Regional Cooperation (SAARC) nations. They reported reasonable levels of accuracy, however, they did not mention any uncertainty measures. Yang et al. (2020) used a dynamic SEIR model, utilising an artificial intelligence (AI) model that was trained on the older SARs data in China. According to the authors, both models were effective in predicting the epidemic peak and size but no uncertainty measures were provided. Doornik et al. (2020) compared the performance of epidemiological models prepared from the Medical Research Council Centre for Global Infectious Disease Analysis at Imperial College London to their own statistical forecasts and discussed the large forecast errors of epidemiological models in the first months of the pandemic.

A model used in the ensemble of the Centre for Disease Control and Prevention (CDC) is the metapopulation SEIR model developed by researchers at the Mailman School of Public Health of Columbia University (Yamana et al., 2020). This model forecasts daily reported cases, daily new infections, bed demand, ICU and ventilator demand, as well as mortality in US counties for six weeks into the future. The Institute of Health Metrics (IHME) created a dashboard for forecasting deaths and hospital resource use for most countries, but their model validation metrics have not yet been published (IHME, 2020; IHME COVID-19 health service utilisation forecasting team & Murray, 2020). The results of this model are also used in the CDC ensemble.

The Institute of Global Health, Faculty of Medicine, University of Geneva and the Swiss Data Science Centre have created a web dashboard that provides forecasts of the number of cases and deaths for the next seven days based on exponential and linear statistical models fit to the recent growth rate of cumulative deaths (Institute of Global Health, Faculty of Medicine, University of Geneva and the Swiss Data Science Center EZ-E, 2020). The University of Texas also created a web dashboard for the projection of deaths in US states for the next seven days using Nonlinear Bayesian hierarchical regression with a negative-binomial model for daily variation in death rates (Woody et al., 2020). Their uncertainty measures have not yet been published. Youyang Gu (2020), an independent data scientist, also produced COVID-19 projections. His model is based on the SEIR model, but he then used machine learning to learn and minimise uncertainty. Uncertainty metrics have not been published yet, however, the model significantly under-forecasted deaths in early and mid-April.

2.2 Forecasting models: Global coverage

A study performed by Elmousalami and Hassanien (2020) attempted to forecast global infected cases, global recovered cases and global deaths. Single Exponential Smoothing was superior to moving average and weighted moving average methods based on several error measures. Another model is the one developed by the Imperial College London (Flaxman et al., 2020; Imperial College COVID-19 response team, 2020). They used an ensemble of transmission approaches to come up with forecasts of deaths for

the following week. Although a detailed explanation of their methodology is provided, the accuracy and bias of the forecasts have not yet been published.

Auquan (2020) have created a COVID-19 dashboard that uses a modified SEIR model to provide 10-day-ahead forecasts both globally and country-specific. Other than model structure, no other specifics are available. This is one of the models that the Centre for Disease Prevention (CDC) is using within their ensemble of forecasts. Similarly, the DELPHI model forecasts infections, hospitalisations, and deaths for the next 45 days and was developed by the MIT team of Operations Research and Analytics (2020). The model is an adjusted SEIR model with added features such as underdetection and differentiated government intervention. This model's results are also used by the CDC ensemble model but their detailed methodology and forecast's efficiency have not yet been reported.

Good Judgment Inc, a consultancy using judgmental forecasting methodology to help clients make better decisions, created a COVID dashboard which forecasts notable socioeconomic milestones regarding COVID-19, such as the development of a vaccine and return to normal air travel (Good Judgment, 2020). They ask specific questions and, by utilising the wisdom of their trained crowd, manage to assign a probability to the occurrence of an event, enabling informed decisions. The accuracy of the forecasts using this method is based on the forecaster's subject matter knowledge.

3. Method and Analysis

3.1 Data, set-up, and model

Our model predicts two variables related to COVID-19, namely the cumulative number of confirmed cases and cumulative number of deaths. Our core level of focus is global, however, we also extend our forecasts to a small set of countries that responded differently with regards to addressing the pandemic. We are interested in both point-forecast accuracy and prediction interval performance for three confidence intervals: 50, 70, and 90%.

We retrieved data per country from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (2020). This source provides confirmed cases, deaths, and recovered cases per country (and in some cases per state) since 2020-01-22¹, which were then aggregated to a global level. To take into account revisions in the data, from 2020-04-13 onwards we started creating parallel databases while keeping the previous set of data to allow the reproduction of our forecasts. While data revisions resulted in some cases to slight changes in the past forecasts and prediction intervals, these changes were not significant. In particular, there were no differences in the cases that we were over or under forecasting.

We followed a rolling-origin evaluation process (Tashman, 2000). We started with 10 data points available (from 2020-01-22 to 2020-01-31) and produced forecasts and prediction intervals for the next 10 days

¹ We are using the ISO 8601 format to express dates: YYYY-MM-DD.

(2020-02-01 to 2020-02-10). On 2020-02-11, we expanded the in-sample sets to include 20 data points (from 2020-01-22 to 2020-02-10) and produced, again, forecasts for the next 10 days. This process was repeated another 10 times. In total, we produced 12 rounds of 10-step-ahead non-overlapping forecasts covering a four-month period from February to May 2020. Our choice on the horizon ($h = 1 \dots 10$ days) is in-line with how governments are preparing their short-term planning to respond to this pandemic. For example, in producing their modelling scenarios and forecasting the short-term trajectory of the pandemic, the Public Health Agency of Canada (2020) also uses a 10-day horizon.

We would like to emphasise that, with regards to the confirmed cases, this is not an ex post evaluation but a real-time forecasting exercise that has been taking place since the beginning of this pandemic. The first two authors had been publishing forecasts and updates every 10 days through social media (please, refer to their Twitter accounts, @fotpetr and @spyrosmakrid). While the forecasts for deaths had been produced ex-post for the first seven rounds (up until 2020-04-10), we provided real-time forecasts for this variable as well from mid-April until late May. All other forecasts presented in this paper have been prepared ex-post.

We aim to model the large-scale behaviour of the data (Siegenfeld & Bar-Yam, 2020), avoiding micro-assumptions on a large number of unknown variables (such as transmissibility or death rates). We resort to the exponential smoothing family of models (Gardner, 2006; Hyndman et al., 2002, 2008; Taylor, 2003), which are suitable in capturing and extrapolating the levels, trends, and seasonal patterns in the data. As our data do not contain seasonality on an aggregate level, we are focusing on non-seasonal models. We further narrow our focus on exponential smoothing models with multiplicative trends. Multiplicative trend models are not usually considered for demand and supply chain forecasting as they result in explosive forecasts and, by default, are excluded from most statistical software, such as some forecasting package in R. However, they are very suitable in our case as the nature of the data is exponential. Finally, to be able to capture the significant underlying uncertainty, we opt for an exponential smoothing model with a multiplicative error term. While such a choice will generally result in wider prediction intervals compared to a model with an additive error, we prefer to err on the side of caution, i.e., project more uncertainty than less.

Given the above judgmental model selection process (Petropoulos et al., 2018), the only exponential smoothing model that satisfies our criteria is the non-seasonal multiplicative error and multiplicative trend exponential smoothing model usually denoted as ETS(MMN). The mathematical formulation of the model is as follows:

$$\begin{aligned} l_t &= l_{t-1}b_{t-1}(1 + \alpha\epsilon_t), \\ b_t &= b_{t-1}(1 + \beta\epsilon_t), \\ \mu_{t+h} &= l_t b_t^h, \end{aligned}$$

in which l_t is the estimation of the level component at time period t , b_t is the estimation of the trend component for the same time period, α and β are the smoothing parameters for the level and the trend

respectively, h is the forecast horizon, and, finally, μ_t and ϵ_t are the mean estimate (point forecast) and the respective error for period t . We would like to highlight the multiplicative interactions of the level, trend, and error terms.

We applied the above model using the `ets()` function of the *forecast* package (R. Hyndman et al., 2020) for R statistical software (R version 3.6.0, *forecast* version 8.11) through the RStudio Cloud service. It is important to note that ETS(MMN) lacks analytical expressions, so the estimation of the prediction intervals is done via simulation and assuming a Gaussian distribution. The model parameters were estimated via likelihood maximisation, while a new set of parameters was re-estimated for each round of forecasts as more data became available. In the first seven rounds of forecasts for the confirmed cases, the default parameter restrictions of the `ets()` function were used. We then slightly restricted the lower limit for the smoothing parameters so that , as we noticed that lower values can result in tight prediction intervals and underestimation of the uncertainty. This restriction was applied in the last five rounds of the global confirmed cases (from 2020-04-11 to 2020-05-30) as well as for all rounds of other variables presented in this paper.

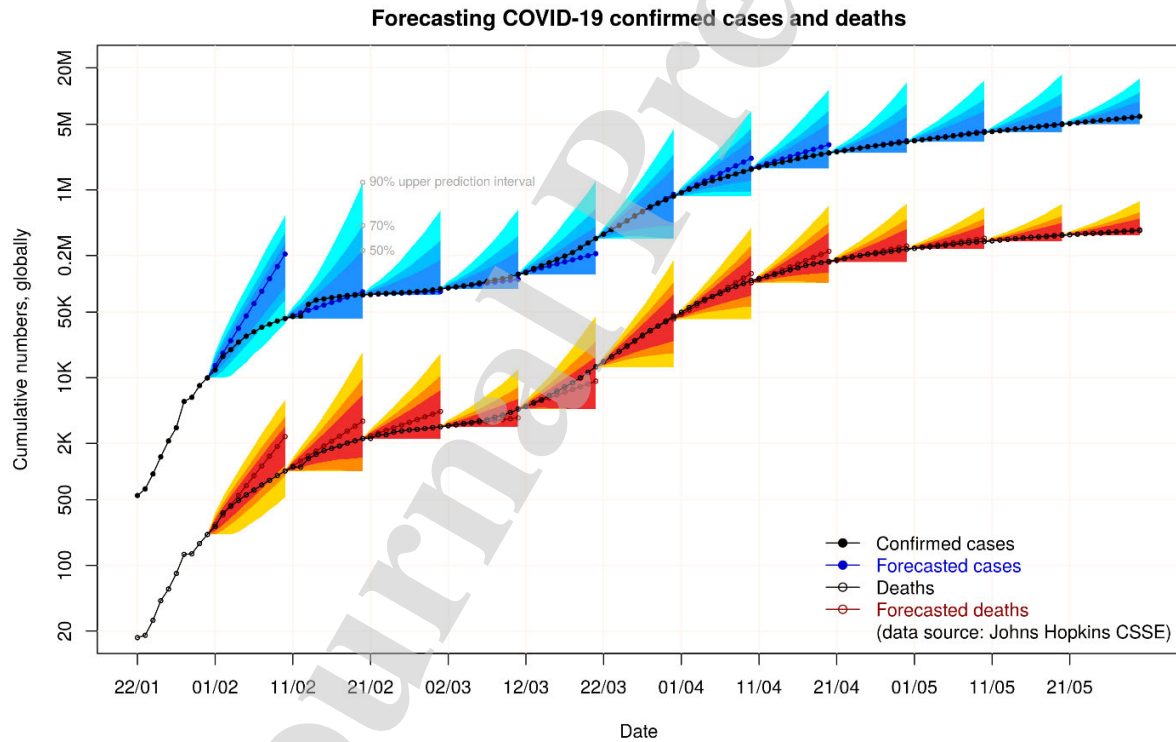


Figure 1. Forecasts and prediction intervals for the global confirmed cases and deaths.

3.2 Global forecasts and their accuracy

Figure 1 depicts the point forecasts and the 50%, 70%, and 90% prediction intervals for confirmed COVID-19 cases (in blue colour) and deaths (in red colour). The vertical axis is log-scaled. The first observation

from the figure is that there is a very strong association between the actual confirmed cases and deaths. As a result, we believe that it is informative to use the former to predict the trajectory of a pandemic. A second observation with regards to the direction of the errors for these two variables is that in all rounds but one (the third round), the sign of the forecast error for the furthest horizon of the confirmed cases and deaths agrees. In other words, if we are over-forecasting one variable then it is very likely that we will also over-forecast the other variable. A third observation has to do with the prediction intervals. The ETS (MMN) results in relatively wide intervals; in all but one of the 12 rounds, both variables lie comfortably within the 50% prediction intervals. While from a calibration point-of-view our forecasts are not ideal, wide prediction intervals are acceptable in low-predictability situations. In particular, we underforecasted the confirmed cases and the deaths in only three and two rounds, respectively, when the virus started picking up outside mainland China. This does not necessarily mean that our produced forecasts were positively biased, but that preventive actions were taken to limit the impact of the pandemic and that such actions resulted in changing the established patterns in the data.

Tables 1 and 2 provide more details on forecasts of the confirmed cases and deaths, respectively. In detail, we provide the forecasts for the dates depicted in the second column, which refer to the 10th horizon (the furthest away from the origin day) of each round, together with the actual observation on the same day and the error and two versions of the absolute percentage error (APE), using as denominators the cumulative actuals and the new confirmed cases/deaths respectively. We also provide the expected percentage increase from the last available observation and a measure for uncertainty (last two columns of tables 1 and 2). The uncertainty is measured as the percentage of the difference between the point forecast and the 50% upper prediction interval in the furthest horizon, divided by the point forecast. This allows us to compare the uncertainty levels across different horizons, given the cumulative nature of the data, i.e., lower values of uncertainty suggest relatively tighter prediction intervals.

Table 1. Actual values, 10-days-ahead forecasts, errors, expected increase and uncertainty for the global **confirmed cases**.

Round	Date	Actual (1000s)	Forecast (1000s)	Error (1000s)	APE (%)	APE on change (%)	Expected increase (%)	Uncertainty (%)
1	2020-02-10	42.8	207.4	-164.6	384.8	501.2	1987	29
2	2020-02-20	76.2	82.0	-5.8	7.6	17.4	92	177
3	2020-03-01	88.4	82.9	5.5	6.2	45.1	9	115
4	2020-03-11	126.9	111.7	15.2	12.0	39.4	26	98
5	2020-03-21	304.4	209.4	95.0	31.2	53.5	65	95
6	2020-03-31	857.5	900.5	-43.1	5.0	7.8	196	89
7	2020-04-10	1691.7	2173.2	-481.5	28.5	57.7	153	58
8	2020-04-20	2472.3	3032.4	-560.1	22.7	71.8	79	69
9	2020-04-30	3256.8	3343.9	-87.0	2.7	11.1	35	78
10	2020-05-10	4101.7	4209.8	-108.1	2.6	12.8	29	62
11	2020-05-20	4996.5	5067.9	-71.4	1.4	8.0	24	58
12	2020-05-30	6059.0	6068.4	-9.4	0.2	0.9	21	45

Several observations arise from tables 1 and 2:

- A large forecast error is associated with changes in the observed patterns. With regards to the confirmed cases, this is true in rounds 1 (where a significant decline is observed in China), in 4-5 (associated with the spread in other countries), and in 7-8 (associated with global measures for controlling the spread, such as lockdowns). In all other rounds, the percentage of forecast error for confirmed cases is single digit. In particular, the error for the last four rounds is lower than 3%.
- While the signed error and the APE can inform us of what happened and whether the applied policies and measures have been successful, the expected increase can inform decisions made with regards to retaining, strengthening or relaxing such measures. We observe a decrease in the expected increase rate in the last rounds, but still, the trend is substantial: a 24% expected increase in the number of confirmed cases in round 11 suggests an additional 1 million confirmed infections in absolute terms in a period of 10 days.
- Over time, we also observe a decrease in the forecast uncertainty, in terms of the width of the prediction intervals. Additionally, the uncertainty in forecasting deaths is lower than that of the confirmed cases in all rounds. This is to be expected, as the number of confirmed cases heavily depends on the number of tests conducted. While undercounting is to be expected for both variables, it should be much less in the number of deaths.

Table 2. Actual values, 10-days-ahead forecasts, errors, expected increase and uncertainty for the global deaths.

Round	Date (of furthest horizon)	Actual (1000s)	Forecast (1000s)	Error (1000s)	APE (%)	APE on change (%)	Expected increase (%)	Uncertainty (%)
1	2020-02-10	1.01	2.36	-1.34	132.7	168.0	1007	27
2	2020-02-20	2.25	3.44	-1.19	52.9	96.3	239	91
3	2020-03-01	3.00	4.36	-1.37	45.6	182.3	94	69
4	2020-03-11	4.65	3.73	0.91	19.6	55.3	25	59
5	2020-03-21	12.97	9.18	3.79	29.2	45.5	98	81
6	2020-03-31	42.11	45.84	-3.73	8.9	12.8	253	68
7	2020-04-10	102.53	128.59	-26.06	25.4	43.1	205	55
8	2020-04-20	169.99	222.08	-52.10	30.6	77.2	117	58
9	2020-04-30	233.39	252.92	-19.53	8.4	30.8	49	51
10	2020-05-10	282.71	305.71	-23.00	8.1	46.6	31	38
11	2020-05-20	328.12	334.36	-6.25	1.9	13.8	18	34
12	2020-05-30	369.13	376.97	-7.84	2.1	19.1	15	31

Next, we explore the accuracy of our exponential smoothing model ETS (MMN) in forecasting the confirmed cases and deaths for each of the 10 horizons considered. The results are summarised for every four rounds and overall. For both variables, we observe a noticeable decrease in the mean absolute percentage error (MAPE) from the first rounds to the latter ones. The overall forecast error in deaths is generally lower compared to that of the confirmed cases for the same horizons. For example, the five-step-ahead MAPE is 9.4 and 7.6% for confirmed cases and deaths, respectively. At the same time, the average forecast error for confirmed cases in the latter rounds has been low, below 2% at the furthest horizon.

Table 3. Point-forecast accuracy (MAPE, %) for confirmed cases per horizon.

Rounds	Horizon									
	1	2	3	4	5	6	7	8	9	10
1 to 4	3.6	4.6	9.8	14.7	20.5	29.1	39.5	54.5	73.9	102.6
5 to 8	2.2	2.2	3.9	5.0	7.2	9.4	11.6	14.6	18.3	21.8
9 to 12	0.2	0.3	0.3	0.3	0.4	0.6	1.0	1.2	1.5	1.7
Average	2.0	2.4	4.7	6.7	9.4	13.0	17.4	23.4	31.2	42.1

Table 4. Point-forecast accuracy (MAPE, %) for **deaths** per horizon.

Rounds	Horizon									
	1	2	3	4	5	6	7	8	9	10
1 to 4	3.8	6.5	5.6	8.9	14.2	21.6	30.6	39.7	50.4	62.7
5 to 8	2.1	2.0	3.0	4.9	7.1	9.2	10.9	14.9	19.2	23.5
9 to 12	0.7	0.8	0.9	1.1	1.5	2.2	3.0	3.6	4.2	5.1
Average	2.2	3.1	3.2	5.0	7.6	11.0	14.8	19.4	24.6	30.5

While our proposed simple model performs best at an aggregate level, it could be equally applied at a country level. For instance, an application of exponential smoothing with multiplicative trend to forecasting the number of COVID-19 deaths for the United States results in an average error of 4.4%, 8.1% and 16.3% for one, two and four-steps-ahead respectively when we focus the evaluation for the rounds 6 to 9 of our application (2020-03-21 to 2020-04-30). Doornik et al. (2020) report average errors of 4.9%, 8.8% and 14.0% for the same lead times and a similar timeframe (2020-03-24 to 2020-04-25, see their table 2). Doornik et al. (2020) also mention that their approach offers superior performance against two epidemiological models. This suggests that our simple approach performs well in terms of accuracy against other published forecasts.

3.3 An examination of countries with diverse government responses

We now compare the death forecasts for three Scandinavian countries, namely Sweden, Denmark and Norway. The Oxford COVID-19 Government Response Tracker (University of Oxford, 2020) tracks the stringency level of each country, which records the strictness of country-level policies with regards to the containment of the virus, such as lockdowns. The stringency level for these three countries on 2020-03-30 was 38.22, 84.12 and 79.63 (out of 100), with higher values suggesting stricter measures². For example, the closure of schools at all levels and the cancellation of public events was required in Denmark and Norway; at the same time, the schools' closure was only a recommendation in Norway, with no measures taken on cancelling public events. Denmark and Norway recommended the closure of public transport; Sweden took no similar measures. Denmark and Norway recommended and required restrictions on internal movement; Sweden did not. All three countries have had coordinated public campaigns and similar testing policies (symptomatic & eligible).

² The data were collected on 2020-05-16.

In figure 2, we present the actual values of deaths per million in each of these three countries, along with the 10-step-ahead forecasts and the 70% prediction intervals. Note that we started producing forecasts once 10 or more non-zero observations became available for each country. We observe that the differences in measures were translated into significant differences in the number of deaths per million as well as the respective forecasts.

While Sweden and Denmark were at similar levels at the end of March, the additional measures taken by the latter decreased the rate of deaths more effectively compared to the former. A high stringency index is also associated with a decreased forecast uncertainty, both in absolute and relative terms. Whether or not Sweden's herd immunity plan was successful can only be judged in the long term but the results until now have not been encouraging (Tangermann, 2020). Comparing Denmark with Norway, the former applied stricter policies and measures, such as recommendations to "stay at home". This is reflected in the difference in the stringency index between the two countries. Regardless, Norway had fewer deaths compared to Denmark. This can be attributed to factors out of the governments' control, such as population density (13.9/km² for Norway versus 135.7/km² for Denmark) and the percentage of the population over 65 years old (17.2% versus 19.6%)³.

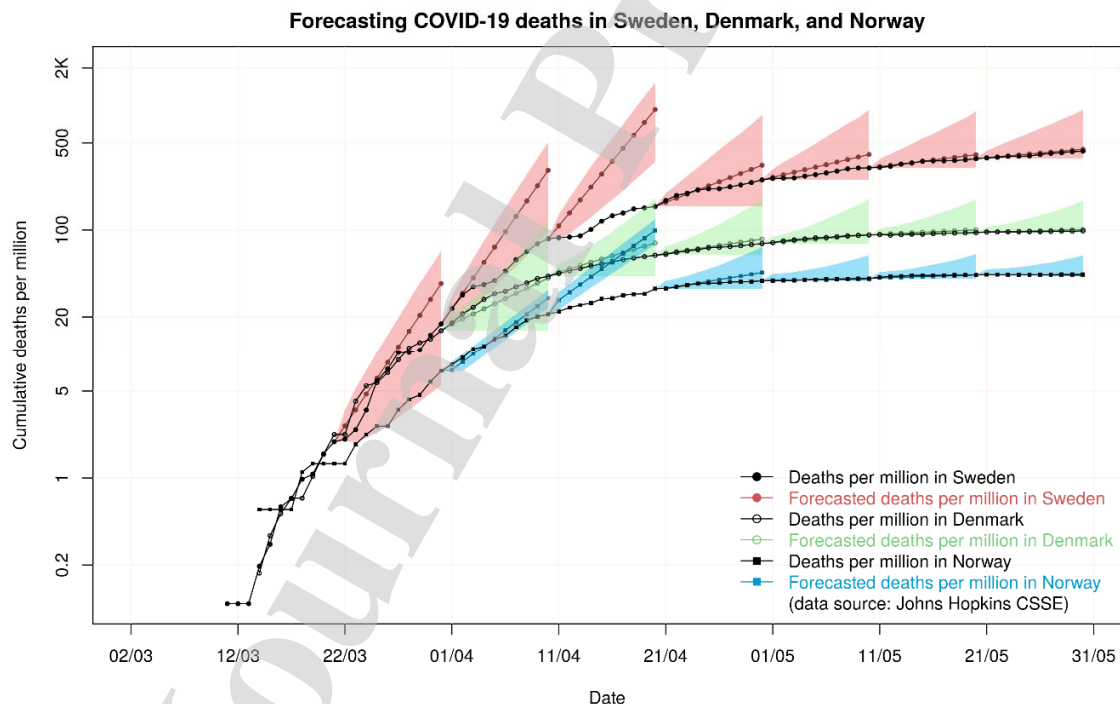


Figure 2. Forecasts and 70% prediction intervals for deaths per million in three countries: Sweden, Denmark, and Norway.

³ <https://www.prb.org/countries-with-the-oldest-populations/> [Accessed on 2020-05-16]

4. Discussion

4.1 Data challenges

Forecasting the trajectory of an early epidemic (before it becomes a pandemic) is challenging, given that limited data points are available and we do not have a good understanding of the transmissibility and death rates. In the absence of data and evidence, a possible sensible approach would be to enforce stronger measures earlier than later. The fat-tailedness of the situation (Cirillo & Taleb, 2020) suggests that underestimating the impact could have catastrophic consequences. At the same time, others suggest that we need to gain a good understanding of the situation (and collect more information and data) to better decide on the appropriate measures to be taken by policy makers (Ioannidis, 2020). In any case, the absence of reliable data reduces our ability to accurately predict the future and take effective actions to minimise the negative consequences of the epidemic/pandemic.

While the quantity of available information is obviously important, its quality and standardisation is arguably even more important. The number of new daily confirmed cases is obviously not equal to the number of new cases, due to limited (and different across countries) testing and the recommendations from health organisations to “stay home” if one has coronavirus-related symptoms. The number of deaths is also under-reported, arguably less than that of confirmed cases. The excess in deaths overall compared to previous years is not consistent with confirmed COVID-19 deaths (Winton Centre for Risk and Evidence Communication, 2020), with the differences being more pronounced for older age groups. Data revisions on the past daily numbers of confirmed cases and deaths have also been an important factor. For instance, on 17th April, China added about 1,300 new deaths for the Wuhan area alone (Yan, 2020). However, data quality issues are expected to differ across countries, with some being better than others in accurately recording and reporting the status quo.

Regardless of the data limitations and inaccuracies, the two variables used in this paper, confirmed cases and deaths, are good indicators for the trajectory of the pandemic and have been consistently used to inform governments in their decision making. “Daily-reported Covid deaths provide an imperfect picture of the effects of the epidemic in any country, both due to reporting delays and the fact that they usually only feature deaths in hospital. Nevertheless, the broad patterns can show the basic trends over time” (Winton Centre for Risk & Evidence Communication, 2020). It is not unusual in the forecasting literature to use proxies for estimating an outcome. An analogy is supply chain forecasting, where sales are usually used as a proxy for the demand. Moreover, the data quality issues would be less amplified on an aggregate (global) level, which is the focus of this study. A top-level approach offers a smoother, less noisy view which is easier to forecast.

4.2 The performance of forecasts and uncertainty

Producing a set of predictions should not be the end target in a forecasting exercise. It is crucial that the forecasts are consistently compared against the actual values, and that the model (and its parameters) are updated as we roll through the origins. Evaluation and measuring errors is described as one of the four main principles in health forecasting (Soyiri & Reidpath, 2013). Our review of other COVID-19 forecasting models in section 2 suggests that a significant number of researchers do not systematically report the accuracy of their previous predictions, rendering the evaluation of their forecasts difficult.

A useful forecasting approach should not be limited to the provision of point forecasts. In our approach, we report both the mean estimate and three levels of uncertainty. The forecast distribution is particularly important in situations characterised by high uncertainty and can be used towards building extreme scenarios. Arguably, our approach produces prediction intervals that are relatively wide. For instance, on 2020-05-21 we predicted that there was a 5% chance that the actual confirmed cases would exceed 15 million when the confirmed cases at the end of 2020-05-20 were 5 million. While this may seem to be excessive and unrealistic, it is worth pointing out that in three out of the 11 previous rounds (2020-02-01 until 2020-05-20) we observed that the confirmed cases increased by a factor of 2.4 or higher in an interval of 10 days. Regardless, as we discussed in section 3, we believe that it is possibly better to overestimate uncertainty (and be prepared for more extreme scenarios) rather than underestimate it. Underestimation of the uncertainty (and underforecasting) was the case with several of the CDC models in April 2020 (Best & Boice, 2020).

In this study, we used a single time series model for producing COVID-19 forecasts. However, a long-standing result in the forecasting literature is that combinations of accurate and diverse forecasts will improve the performance of the individual base models (Lichtendahl & Winkler, 2020). The CDC has utilised forecast ensembles. The grand majority of the individual models in this ensemble have been epidemiological models (Centers for Disease Control and Prevention, 2020). We believe that the inclusion of statistical models, like our simple model or that of Doornik et al. (2020), as well as judgmental methods, would further increase the accuracy of such ensembles.

4.3 Implications

Healthcare systems are becoming increasingly reliant on predictive analytics to better anticipate demand and optimise resource allocation. Forecasting is a valuable tool for predicting health events and situations, like disease outbreaks, which can potentially create excess demand for health services. Forecasting, when used wisely, can be useful in providing early information to service providers and policy/decision makers so they can intervene and take appropriate actions to manage the expected increased demand. Our analysis suggests that our model is consistent in accurately forecasting (within specified uncertainty levels) the number of cases and deaths throughout the duration of the pandemic we covered. We believe that our

approach could prove to be useful to public health officials in order to monitor the spread of the virus. The results were openly shared on social media platforms for a period of 120 days, therefore, anyone could have access to our forecasts to gain an understanding of the direction of the pandemic and also to use them in an ensemble model.

In times of crisis, such as a pandemic, it is important to provide informed resource allocation in a timely manner. We have demonstrated that a simple, quick, accurate forecasting methodology, which is not computationally challenging, can have robust and meaningful results that can be utilised in a timely manner by the appropriate stakeholder. Simple to explain models are usually more trustworthy, increasing the probability of their usage. In an era of a tsunami of data and a blizzard of models, it is important to be able to provide a robust tool that can be easily understood and consequently used to support decisions made by experts.

Forecasting horizons are important in different ways. In the event of a disease outbreak, the first thing that scientists and decision makers want to know is the disease progression and how the crisis is going to develop. Short-term forecasting, like the one proposed in this study, can be useful in resource allocation and immediate planning of healthcare services as well as interventions such as social isolation and lockdowns. Medium and long term forecasting is more suitable for other planning activities, such as ensuring ventilator availability and vaccine production in the case of the current pandemic.

4.4 Limitations

The simple model that we use to forecast confirmed cases and deaths for COVID-19 has certain limitations. As a pure univariate model, it does not take into account the primary drivers of these two variables, such as governmental actions. Simply, our model extrapolates established patterns in the data, assuming that these patterns are true and will continue to hold in the future. Our forecasts will be negatively biased (over-forecast) when actions are taken to prevent the rise of the cases and deaths, such as enforcing lockdowns. Similarly, our forecasts will be positively biased when new spikes in the cases/deaths are observed, as a result, for example, of the transmission of the virus to new territories or in the event of a second/third wave. Our forecasts will perform best when the established patterns remain stable.

Another way to think of this would be as a tracking signal. If our forecasts show that a rise in cases/deaths is to be expected, then such forecasts should mobilise governments to take actions towards slowing down the transmission rate. In this regard, our forecasts could be a useful tool for decision making; predicting what would happen if nothing changes, by capturing the large-scale behaviour of the data. Such forecasts could be one of the scenarios explored, and in others investigating the expected impact of particular policies. Note that accuracy is not necessarily the objective here. Accurately forecasting an increasing trend of deaths would be much less welcome compared to significantly over-forecasting the same variable as the

result of corrective actions. As such, our forecasts are not (and should not be) the target, but can be used towards planning and decision making.

Another limitation of our approach is that it will perform best on an aggregate level. While our approach will produce reasonable forecasts at a country level (as also discussed in the previous section for US, Denmark, Sweden and Norway), its performance is best when considering aggregated data across many countries that provide an inherent degree of smoothing. An alternative would be to produce forecasts at a low level of aggregation (country level or even regional level) and then aggregate the forecasts using the bottom-up approach, or even produce forecasts at various aggregation levels and reconcile the differences with hierarchical approaches.

A final limitation of our study is with regards to the short forecasting horizons considered. While longer-term forecasts could be appropriate in some settings, still, short-term projections can support decision makers in keeping existing measures in place, taking further precautionary actions to minimise the increase in the transmission rate or, if a slow-down is forecasted, to relax the measures taken. Short-term forecasts can also be used for healthcare management decisions, such as the number of nursing staff, bed management, hospital equipment availability, but also in supporting essential retailers in making decisions towards the application of quotas and avoiding stockouts.

5. Concluding remarks

Forecasting the outcomes of a pandemic is a challenging task with massive potential value to decision and policy makers. Over a period of four months, we published live short-term, global forecasts for two key variables related to COVID-19. To do so, we used a simple time series model suitable for capturing multiplicative trends. Our proposed time series model has shown good levels of accuracy and uncertainty, especially as more data was accumulated. Such forecasts are useful for monitoring the progression of the disease and helping policy makers consider appropriate measures to decrease the negative impacts of the pandemic and to implement actions to strengthen or relax mitigating interventions.

Lessons learnt:

- It is imperative that we accept that the next pandemic will also come as a complete surprise and we must be better prepared to face it, even though it may be several years away.
- More focus must be put on collecting high-quality data that follow standardised definitions supplied by the World Health Organisation.
- Simple forecasting models that focus on the large-scale behaviour of the pandemic can be useful.
- Avoiding making unnecessary micro-assumptions simplifies the forecasting process and makes it more transparent to policy and decision makers.
- Time series models are suitable for short-term forecasting of pandemics if they can capture the exponential patterns in the data and as long as such patterns remain constant.

- More computationally intensive and data-hungry models do not necessarily perform better, while they may not even be applicable at the very early stages of a pandemic.
- Forecasting performance should be an integral part of publishing and communicating predictions.

Could future research improve the forecasting of pandemics? Perhaps some improvements are possible by developing new methods that combine forecasts from many diverse sources, including time-series and epidemiological models and using more and higher quality data. Moreover, the performance of models could be enhanced utilising higher quality and more disaggregated levels of data (country and local levels as well as age-specific), possibly through using hierarchical structures and longer forecast horizons. Finally, models could also be adjusted to forecast more outbreak-specific variables, such as demand for ventilators and intensive care unit beds. Although many mistakes were made in dealing with COVID-19, there is a lot to learn from the experience gained so far to help humanity prepare for future pandemics.

Other critical questions for future investigation relate to the long-term conduct of the pandemic and its human and economic impact. Will there be a second and/or third wave? How many people will become infected? What percentage of the infected will die? What will happen as the lockdowns are removed, and social distancing is relaxed? Will the spread of the virus increase? Will the lockdowns and social distancing lead to a major economic recession or even a depression? How will the stock market react, after first ignoring the pandemic until 2nd March when it achieved an all-time high (in the USA), then lost more than a third of its capitalisation 24 days later, and then almost reached its previous all-time height at the beginning of June? What will happen to the 60 million people being pushed into "extreme poverty" by the effects of the coronavirus (David, 2020)? Will the pandemic end without a vaccine and how long will it take until such a vaccine becomes available?

References

- Auquan. (2020). *COVID-19 Dashboard*. <https://covid19-infection-model.auquan.com/>
- Becker, A. D., & Grenfell, B. T. (2017). tsIR: An R package for time-series Susceptible-Infected-Recovered models of epidemics. *PloS One*, 12(9), e0185528.
- Best, R., & Boice, J. (2020). *Where The Latest COVID-19 Models Think We're Headed — And Why They Disagree*. FiveThirtyEight. https://projects.fivethirtyeight.com/covid-forecasts/?ex_cid=rrpromo
- Bootsma, M. C. J., & Ferguson, N. M. (2007). The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18), 7588–7593.
- Brooks, L. C., Farrow, D. C., Hyun, S., Tibshirani, R. J., & Rosenfeld, R. (2015). Flexible Modeling of Epidemics with an Empirical Bayes Framework. *PLoS Computational Biology*, 11(8), e1004382.
- Carras, C. (2020, April 13). *In 2015, Bill Gates predicted an epidemic would kill millions. Here's what he says now*. Los Angeles Times. <https://www.latimes.com/entertainment-arts/story/2020-04-13/coronavirus-bill-gates-ellen-degeneres-ted-talk>
- Centers for Disease Control and Prevention. (2020, May 28). *COVID-19 Forecasts: Cumulative Deaths*. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html#anchor_1587397564229
- Chretien, J.-P., Riley, S., & George, D. B. (2015). Mathematical modeling of the West Africa Ebola epidemic. *eLife*, 4. <https://doi.org/10.7554/eLife.09186>
- Cirillo, P., & Taleb, N. N. (2020). Tail risk of contagious diseases. *Nature Physics*, 1–8.
- Dahir, A. L. (2020, April 22). *"Instead of Coronavirus, the Hunger Will Kill Us." A Global Food Crisis Looms*. New York Times. <https://www.nytimes.com/2020/04/22/world/africa/coronavirus-hunger-crisis.html?referringSource=articleShare>
- David, D. (2020, June 7). *Coronavirus "a devastating blow for world economy."* BBC. <https://www.bbc.co.uk/news/business-52939846>
- Doomik, J. A., Castle, J. L., & Hendry, D. F. (2020). Short-term forecasting of the coronavirus pandemic. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2020.09.003>
- Elmousalami, H. H., & Hassanien, A. E. (2020). Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modeling and Recommendations. In *arXiv [q-bio.PE]*. arXiv. <http://arxiv.org/abs/2003.07778>
- Eyewitness News. (2020, May 5). *Coronavirus News: Governor Cuomo on reopening, "How much is a human life worth?"* ABC7NY. <https://abc7ny.com/coronavirus-new-york-ny-cases-in-news/6153317/>
- Farrow, D. C., Brooks, L. C., Hyun, S., Tibshirani, R. J., Burke, D. S., & Rosenfeld, R. (2017). A human judgment

- approach to epidemiological forecasting. *PLoS Computational Biology*, 13(3), e1005248.
- Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá Perez, Z., Cuomo-Dannenburg, G., & Others. (2020). *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*. <https://doi.org/10.25561/77482>
- Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Coupland, H., Mellan, T. A., Zhu, H., Berah, T., Eaton, J. W., Guzman, P. N. P., Schmit, N., Callizo, L., Imperial College COVID-19 Response Team, Whittaker, C., Winskill, P., Xi, X., Ghani, A., Donnelly, C. A., Riley, S., ... Bhatt, S. (2020). Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in European countries: technical description update. In *arXiv [stat.AP]*. arXiv. <http://arxiv.org/abs/2004.11342>
- Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, 22(4), 637–666.
- Good Judgment. (2020). *COVID Recovery Dashboard*. Good Judgment. <https://goodjudgment.com/covidrecovery/>
- Gupta, R., & Pal, S. K. (2020). Trend Analysis and Forecasting of COVID-19 outbreak in India. *medRxiv*, 2020.03.26.20044511.
- Gu, Y. (2020). *COVID-19 Projections Using Machine Learning*. <https://covid19-projections.com/>
- Hyder, A., Buckeridge, D. L., & Leung, B. (2013). Predictive validation of an influenza spread model. *PloS One*, 8(6), e65459.
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., & Yasmeen, F. (2020). *forecast: Forecasting functions for time series and linear models*. <http://pkg.robjhyndman.com/forecast>
- Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Verlag.
- Hyndman, R. J., Koehler, A. B., Snyder, R., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.
- IHME. (2020). *COVID-19 Projections*. COVID-19 Health Data. <https://covid19.healthdata.org/united-states-of-america>
- IHME COVID-19 health service utilization forecasting team, & Murray, C. J. L. (2020). 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. In *Infectious Diseases (except HIV/AIDS)* (No. medrxiv;2020.04.21.20074732v1). medRxiv. <https://doi.org/10.1101/2020.04.21.20074732>
- Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P., & Hashizume, M. (2015). Time series regression model for infectious disease and weather. *Environmental Research*, 142, 319–327.

Imperial College COVID-19 response team. (2020). *Short-term forecasts of COVID-19 deaths in multiple countries*.

<https://mrc-ide.github.io/covid19-short-term-forecasts/index.html#individual-model-outputs>

Institute of Global Health, Faculty of Medicine, University of Geneva and the Swiss Data Science Center EZ-E.

(2020). *COVID-19 Daily Epidemic Forecasting*. <https://renkulab.shinyapps.io/COVID-19-Epidemic-Forecasting/>

Ioannidis, J. P. A. (2020, March 17). *A fiasco in the making? As the coronavirus pandemic takes hold, we are making decisions without reliable data*. STAT. <https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/>

Johns Hopkins University Center for Systems Science and Engineering. (2020). *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*. GitHub.

<https://github.com/CSSEGISandData/COVID-19>

Kandula, S., Yamana, T., Pei, S., Yang, W., Morita, H., & Shaman, J. (2018). Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. *Journal of the Royal Society, Interface / the Royal Society*, 15(144), 20180174.

Kermack, W. O., McKendrick, A. G., & Walker, G. T. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772), 700–721.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., & Surendran, S. (1999). Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble. *Science*, 285(5433), 1548–1550.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32(2), 311–328.

Lichtendahl, K. C., & Winkler, R. L. (2020). Why do some combinations perform better than others? *International Journal of Forecasting*, 36(1), 142–149.

Lilla, M. (2020, May 22). *No One Knows What's Going to Happen*. New York Times.

<https://www.nytimes.com/2020/05/22/opinion/sunday/coronavirus-prediction-future.html>

Li, Q., Guo, N.-N., Han, Z.-Y., Zhang, Y.-B., Qi, S.-X., Xu, Y.-G., Wei, Y.-M., Han, X., & Liu, Y.-Y. (2012). Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic fever with renal syndrome. *The American Journal of Tropical Medicine and Hygiene*, 87(2), 364–370.

Mahbubani, R. (2020, May 30). *Hundreds of schools in South Korea reopened, only to close again as the country sought to avoid a spike in coronavirus cases*. Business Insider. <https://www.businessinsider.com/hundreds-south-korean-schools-closed-amid-spike-coronavirus-cases-2020-5?r=US&IR=T>

MIT Operations Research Center. (2020). *DELPHI Epidemiological Case Predictions*. COVID Analytics.

<https://www.covidanalytics.io/projections>

Osterholm, M. T. (2005, July 1). *Preparing for the Next Pandemic*. Foreign Affairs.

<https://www.foreignaffairs.com/articles/2005-07-01/preparing-next-pandemic>

Osterholm, M. T., & Olshaker, M. (2020, May 21). *Chronicle of a Pandemic Foretold*. Foreign Affairs.

<https://www.foreignaffairs.com/articles/united-states/2020-05-21/coronavirus-chronicle-pandemic-foretold>

Pei, S., Kandula, S., & Shaman, J. (2020). Differential Effects of Intervention Timing on COVID-19 Spread in the United States. In *Infectious Diseases (except HIV/AIDS)* (No. medrxiv;2020.05.15.20103655v1). medRxiv.

<https://doi.org/10.1101/2020.05.15.20103655>

Petropoulos, F., Kourntzes, N., Nikolopoulos, K., & Siemsen, E. (2018). Judgmental selection of forecasting models. *Journal of Operations Management*, 60, 34–46.

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PloS One*, 15(3), e0231236.

Philemon, M. D., Ismail, Z., & Dare, J. (2019). A Review of Epidemic Forecasting Using Artificial Neural Networks. *International Journal of Epidemiologic Research*, 6(3), 132–143.

Piccolomini, E. L., & Zama, F. (2020). Monitoring Italian COVID-19 spread by a forced SEIRD model. *PloS One*, 15(8), e0237417.

Public Health Agency of Canada. (2020, May 3). *COVID-19 in Canada: Using data and modelling to inform public health action*. Government of Canada. <https://www.canada.ca/en/public-health/services/publications/diseases-conditions/covid-19-using-data-modelling-inform-public-health-action.html>

Rogers, A. (2020, May 11). *How Much Is a Human Life Actually Worth?* Wired. <https://www.wired.com/story/how-much-is-human-life-worth-in-dollars/>

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), 20425–20430.

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012-2013 season. *Nature Communications*, 4, 2837.

Siegenfeld, A. F., & Bar-Yam, Y. (2020). What models can and cannot tell us about COVID-19. *Working Paper*. <https://static1.squarespace.com/static/5b68a4e4a2772c2a206180a1/t/5e8e5ac9e696ba4b972093fa/1586387658680/Modeling2.pdf>

Smith, M. E., Singh, B. K., Irvine, M. A., Stolk, W. A., Subramanian, S., Hollingsworth, T. D., & Michael, E. (2017). Predicting lymphatic filariasis transmission and elimination dynamics using a multi-model ensemble framework. *Epidemics*, 18, 16–28.

Soyiri, I. N., & Reidpath, D. D. (2013). An overview of health forecasting. *Environmental Health and Preventive*

- Medicine*, 18(1), 1–9.
- Spaeder, M. C., & Fackler, J. C. (2012). A multi-tiered time-series modelling approach to forecasting respiratory syncytial virus incidence at the local level. *Epidemiology and Infection*, 140(4), 602–607.
- Statista. (2020). *Coronavirus (COVID-19) deaths worldwide per one million population as of May 29, 2020, by country*. <https://www.statista.com/statistics/1104709/coronavirus-deaths-worldwide-per-million-inhabitants/>
- Sujath, R., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment: Research Journal*. <https://doi.org/10.1007/s00477-020-01827-8>
- Tabataba, F. S., Chakraborty, P., Ramakrishnan, N., Venkatramanan, S., Chen, J., Lewis, B., & Marathe, M. (2017). A framework for evaluating epidemic forecasts. *BMC Infectious Diseases*, 17(1), 345.
- Taleb, N. N. (2008). *The Black Swan: The Impact of the Highly Improbable* (New Edition). Penguin.
- Tangermann, V. (2020, June 3). *Sweden's hands-off approach to COVID has failed dramatically*. Futurism. <https://futurism.com/neoscope/swedish-officials-regret-death-toll>
- Tapak, L., Hamidi, O., Fathian, M., & Karami, M. (2019). Comparative evaluation of time series models for predicting influenza outbreaks: application of influenza-like illness data from sentinel sites of healthcare centers in Iran. *BMC Research Notes*, 12(1), 353.
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), 437–450.
- Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: the mother of all pandemics. *Emerging Infectious Diseases*, 12(1), 15–22.
- Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19(4), 715–725.
- Tseng, Y.-J., & Shih, Y.-L. (2019). Developing epidemic forecasting models to assist disease surveillance for influenza with electronic health records. *International Journal of Computers & Applications*, 1–6.
- University of Oxford. (2020). *Oxford COVID-19 Government Response Tracker*. <https://covidtracker.bsg.ox.ac.uk/>
- Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., Vespignani, A., & RAPIDD Ebola Forecasting Challenge group. (2018). The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22, 13–21.
- Walsh, B. (2017, May 4). *The World Is Not Ready for the Next Pandemic*. Time. <https://time.com/4766624/next-global-security/>
- Wang, Z., Chakraborty, P., Mekaru, S. R., Brownstein, J. S., Ye, J., & Ramakrishnan, N. (2015). Dynamic poisson

autoregression for influenza-like-illness case count prediction. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1285–1294.

Winton Centre for Risk and Evidence Communication. (2020). *Winton Centre for Risk and Evidence Communication*. University of Cambridge. <https://wintoncentre.maths.cam.ac.uk/>

Winton Centre for Risk & Evidence Communication. (2020). *Covid-19 Chart*. University of Cambridge. <https://wintoncentre.maths.cam.ac.uk/coronavirus/chart/>

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*, 2020.04.16.20068163.

Yamana, T., Pei, S., Kandula, S., & Shaman, J. (2020). Projection of COVID-19 Cases and Deaths in the US as Individual States Re-open May 4,2020. *medRxiv*, 2020.05.04.20090670.

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., ... He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, 12(3), 165–174.

Yan, S. (2020, April 17). *China adds nearly 1,300 coronavirus deaths to official Wuhan toll, blaming reporting delays*. Telegraph. <https://www.telegraph.co.uk/news/2020/04/17/china-adds-nearly-1300-coronavirus-deaths-official-wuhan-toll/>

Figure 1

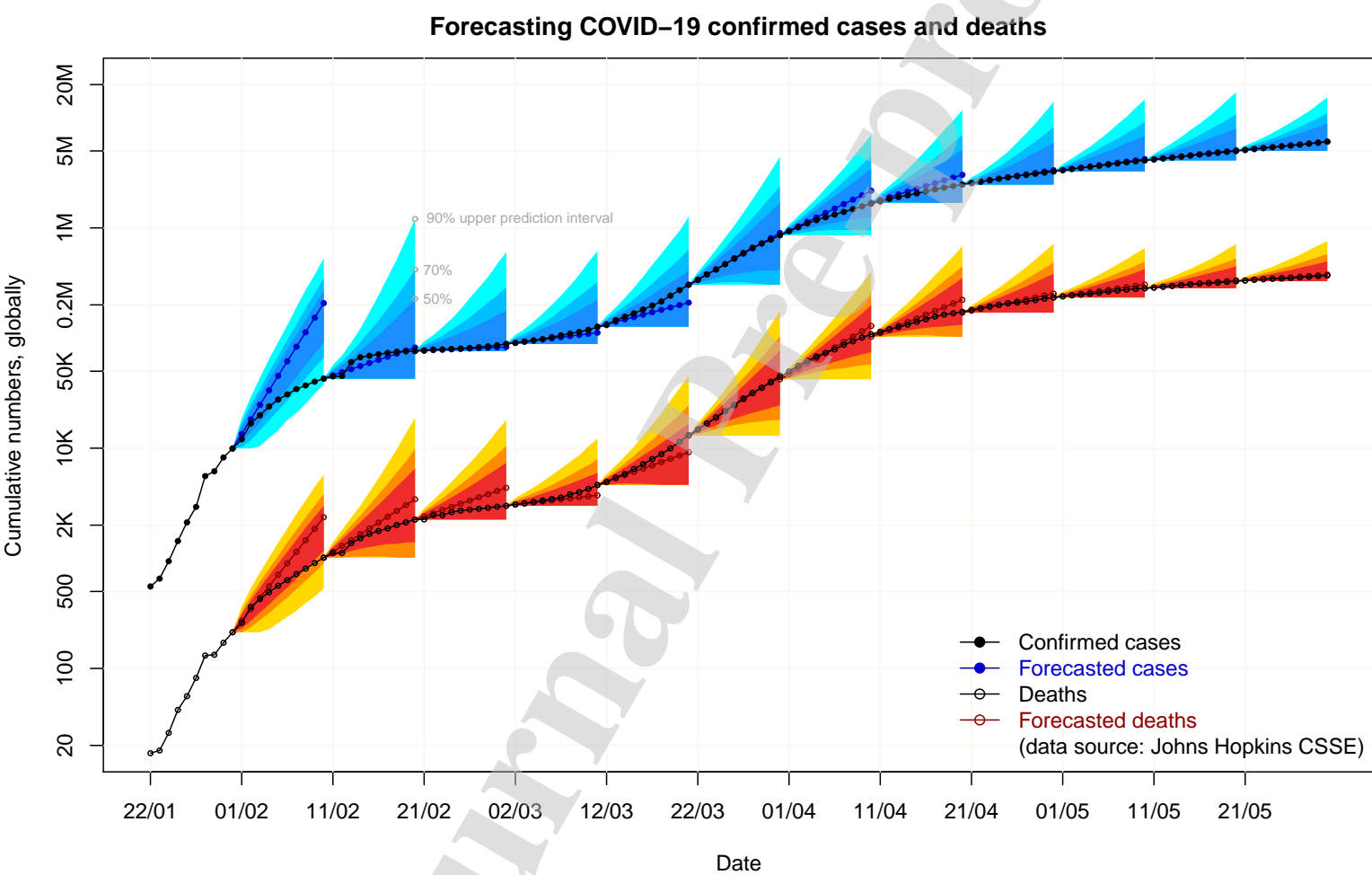
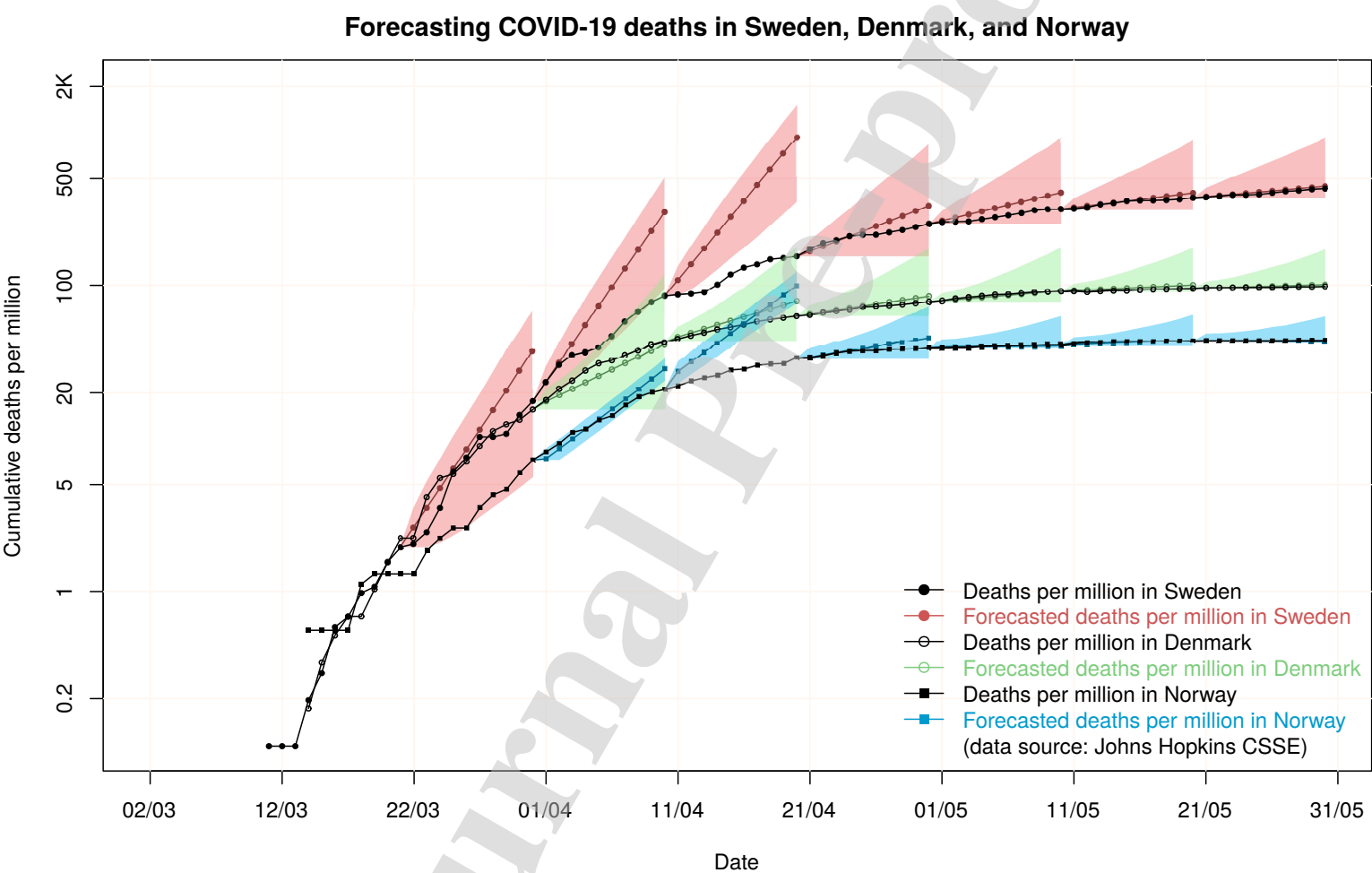


Figure 2



Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

--