Guiding Austria through the COVID-19 Epidemics with a Forecast-Based Early Warning System

Martin Bicher[1,2,†], Martin Zuba[3,†], Lukas Rainer[3], Florian Bachner[3], Claire Rippinger[2], Herwig Ostermann[3,4], Nikolas Popper[1,2,5], Stefan Thurner[6,7,8,9], Peter Klimek[6,7*]

[1] *Institute of Information Systems Engineering, TU Wien, Favoritenstraße 8-11, A-1040 Vienna, Austria*

[2] *dwh simulation services, dwh GmbH, Neustiftgasse 57-59, A-1070 Vienna, Austria*

[3] *Austrian National Public Health Institute, Stubenring 6, A-1010 Vienna, Austria*

[4] *Private University for Health Sciences, Medical Informatics and Technology GmbH, UMIT, Eduard-Wallnöfer-Zentrum 1, A-6060 Hall in Tirol, Austria*

[5] *Society for Decision Support Policy and Planning, DEXHELPP, Neustiftgasse 57-59, A-1070 Vienna, Austria*

[6] *Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria*

[7] *Complexity Science Hub Vienna, Josefstädterstraße 39, A-1080 Vienna, Austria*

[8] *International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361 Laxenburg, Austria*

[9] *Santa Fe Institute, 1399 Hyde Park road, Santa Fe, NM 87501, USA*

[†] *equal contributions*

(Dated: July 2020)

Background: The corona crisis hit Austria at the end of February 2020 with one of the first European superspreading events. In response, the governmental crisis unit commissioned a forecast consortium with regularly projections of case numbers and demand for hospital beds.

Methods: We consolidated the output of three independent epidemiological models (ranging from agent-based micro simulation to parsimonious compartmental models) and published weekly short-term forecasts for the number of confirmed cases as well as estimates and upper bounds for the required hospital beds.

Findings: Here, we report om four key contributions by which our forecasting and reporting system has helped shaping Austria's policy to navigate the crisis and re-open the country step-wise, namely (i) when and where case numbers are expected to peak during the first wave, (ii) how to safely re-open the country after passing this peak, (iii) how to evaluate the effects of non-pharmaceutical interventions and (iv) provide hospital managers guidance to plan health-care capacities.

Interpretation. Complex mathematical epidemiological models play an important role in guiding governmental responses during pandemic crises, provided they are used as a monitoring system to detect epidemiological change points. For policy-makers, the media and the public, it might be problematic to distinguish short-term forecasts from worst-case scenarios with undefined levels of certainty, creating distrust in the legitimacy and accuracy of such models. However, when used as a short-term forecast-based monitoring system, the models can inform decisions to ease or strengthen governmental responses.

I. INTRODUCTION

The first known COVID-19 cases in Austria appeared at the end of February 2020 together with one of the first European superspreading events in the Tyrolean tourist region of Ischgl, visited by travellers from all over the globe [1]. In the first half of March 2020, a nationwide spread of the virus occurred with an exponential rise of confirmed cases [2]. These developments occurred against the dramatic backdrop of the neighboring country of Italy, where despite strict non-pharmaceutical interventions (NPIs) case numbers kept surging, hospital capacities were exceeded and the military had to step in to remove piling bodies [3, 4]. To understand how likely similar developments would have been in Austria, mid-March a forecast consortium was formed and tasked by the government with a weekly forecasting the expected developments in case numbers and how these developments would translate into demand for healthcare re-

* correspondence to: peter.klimek@meduniwien.ac.at

sources. The overarching policy goal at this stage was to navigate the crisis without overburdening the Austrian healthcare system.

At an earlier stage than other middle European countries, Austria took a series of non-pharmaceutical interventions (NPIs) in response to the crisis [5]. Next to a ramping up of healthcare and public health capacities, airport restrictions and landing bans intensified in the first week of March, also targeting other countries than China. Gatherings were limited to 500 persons, cultural and other events started to be cancelled on March 10. On March 16, Austria went into a full lockdown with schools, bars, restaurants, and shops being closed, as well as a transitioning into home office for all non-essential employees [5]. Together with other, earlier measures, these NPIs effectively led to a rapid reduction of daily infection numbers. The number of new cases per day reached a first peak on March 26 with 1,065 cases [6]. COVID-19 related hospitalisations peaked on March 31 with 912 regular beds, whereas the ICU utilization peaked on April 8 with 267 beds being occupied by COVID-19 patients. Daily new cases decreased over April after which they fluctuated at values below one hundred until July [7]. Austrian ICU capacities, estimated to be around 1,000 beds that could have been used for COVID-19 patients while maintaining enough capacity for non-COVID-19 emergencies, have never been in danger of being exceeded in the considered time period [8].

The Austrian COVID-19 forecast consortium provided short-term forecasts for case numbers and required hospital beds. Our consortium consisted of three independent modelling teams with experience in the use and development of sophisticated mathematical and computational models to address epidemiological and public health challenges [9–14]. The consortium was complemented with experts from the Ministry of Health, the Austrian Agency for Health and Food Safety, as well as external public health experts in weekly meetings. A plethora of epidemiological models to forecast the spread of COVID-19 has been proposed recently [15–20]. Here, we consolidated the output of three models into a single forecast of case numbers for 8 days and used these case numbers to predict the numbers of required hospital and ICU beds for 14 days for the country as a whole and for each of its nine federal states over the next two weeks. In addition to these point estimates, we also provided upper and lower bounds for these numbers at various levels of uncertainty. These upper bounds of the hospital bed forecasts served as a guidance system for the regional hospital managers, allowing them to estimate how many beds should be reserved for COVID-19 patients if they were willing to accept a given level of risk. These forecasts have been published each week on the homepage of the Ministry of Health. [21].

At the very beginning of our work as consortium we decided that short-term forecasts have to be clearly separated from long-term scenarios. Due to the multiplicative growth of uncertainties in epidemiological models, accurate forecasts are typically only possible over a time horizon of several days [22–25]. For longer term scenarios that span several weeks, months or even years, however, there is no meaningful way to estimate their uncertainty. For policy-makers and non-technical experts, however, it would not be immediately clear whether a certain projection is a prognosis with a defined level of certainty, or a hypothetical what-if experiment. Therefore, we decided to publish only short-term forecasts.

In this work we present the forecast and reporting system we developed based on the three independent forecasting models. After a brief summary of the individual models, we describe three different strategies we used to combine their outputs, to forecast healthcare demand based on the combined output and to communicate the joint forecast. To evaluate the impact of certain policies (e.g. the lockdown), we report numerical experiments that show how the epidemic would have progressed if measures would have been taken later or not at all. We discuss how our results were received by policy-makers, stakeholders in the healthcare system, and the public. We claim that our approach offered valuable contributions to chart a safe path to re-open the country after the lockdown using a strategy that can be described as "driving on sight". The aim of this work is to communicate the methods applied and developed which allowed three individually thinking modelling and simulation research units to work together in a joint task force producing a consolidated forecast, the benefits and shortcomings of the process, and the political impact of the achieved results. We conclude that epidemiological models can be useful as the basis for short-term forecast-based monitoring systems to detect epidemiological change points, but become problematic when used to produce long-term worst-case scenarios due to their undefined levels of certainty.

II. METHODS

We used three conceptually different epidemiological COVID-19 models, developed and operated individually by three research institutions, namely a modified SIR-X differential equation model (Medical University of Vienna / Complexity Science Hub), an Agent-Based simulation model (TU Wien / dwh GmbH), and a simplified state space model (Austrian National Public Health Institute).

A. Data

Although the three models use different parameters and parametrization routines, they are calibrated using the same data to generate weekly forecasts. Consequently, differences between the model forecasts are a result of different model structure and calibration, but not a result of different data sources. The models also used different nowcasting approaches to correct for late

reporting of positive test results. We used data from the official Austrian COVID-19 disease reporting system (EMS, [7]). The system is operated by the Austrian Ministry of Health, the federal administrations, and the Austrian Agency for Health and Food Safety.

For every person tested positively in Austria, it contains information on the date of the test, date of recovery or death, age, sex and place of residence. Furthermore, hospital occupancy of COVID-19 patients in ICU and normal wards are available from daily reports collected by the Ministry for Internal Affairs.

B. Extended SIR-X Model

One of our models is an extension of the recently introduced SIR-X model [15]. The original SIR-X model introduced a parsimonious way to extend the classic SIR dynamics with the impact of NPIs. In particular, two classes of NPIs are considered. First, there are NPIs that lead to a contact reduction of *all* individuals (susceptible and infected ones). Such NPIs include social distancing and other lockdown measures. Second, the model also represents NPIs that reduce the effective duration of infectiousness for infected individuals. Contact-tracing and quarantine belong to this category.

The original SIR-X model does not offer a way to model the return-to-normal, i.e., the taking back of NPIs. We extended the model by introducing a mechanism by which susceptible but quarantined individuals increase their number of contacts again; a model we dub the XSIR-X model, see SI Appendix A.. Further, we structured the population according to age, introduced multiple calibration phases to model behavioural changes in the population over time, and used mobility data to identify such turning points [26]. Forecast errors are estimated by recalibrating the model to perturbed data points that are displaced proportionally to the empirical deviation between model and data; see SI Appendix A.

C. Agent-Based SEIR Model

The second model is an Agent-Based SEIR type model [12]. It is stochastic, population-dynamic and depicts every inhabitant of Austria as one model agent. It uses sampling methods to generate an initial agent population with statistically representative demographic properties and makes use of a partially event-based, partially time-step (1 day)-based update strategy to enhance in time.

It is based on a validated population model of Austria including demographic processes like death, birth, and migration [14]. Contacts between agents are responsible for disease transmission and are sampled via locations in which agents meet: schools, workplaces, households and leisure-time. After being infected, agents go through a detailed disease and/or patient pathway that depicts the different states of the disease and the treatment of the patient.

The model input consists of a time-line of modelled NPIs; parameters are calibrated using a modified bisection method. Results are gathered via Monte-Carlo simulations as the point-wise sample mean of multiple simulation runs. Due to the large number of agents in the model, 8 simulation runs are used which are sufficient to have the sample mean approximate the real unknown mean with an error of less than 1% with 95% confidence (estimated by the Gaussian stopping introduced in [27]).

The model considers uncertainty with respect to the stochastic perturbations in the model by tracing the standard deviation of the Monte Carlo simulations. Parameter uncertainty is considered in form of manually defined best and worst case scenarios. Parameter values of the model are continuously improved and available online [28].

D. Epidemiological Clockwork Model

The third model is a simplified state space model that traces individuals through the stages infected, latency period, infectious, reported, contained and immunized/deceased; see also SI Appendix B. Based on the ratio between infected and infectious, the true infection rate is calculated and extrapolated in the future using exponential smoothing or moving average.

The model distinguishes between detected an undetected cases and accounts for the import of infected cases (who did not acquire the infection from the calculated number of infectious individuals).

The underlying detection rate, immigration, and the effectiveness of contact-tracing are time-varying parameters that also reflect qualitative information such as spikes in the number of reported new infected cases that can be explained by mass test screenings.

Uncertainty can be modelled by varying underlying model parameters and also results from parameter uncertainty in the infection rate extrapolation.

E. Model harmonization

In order to harmonize the models and generate a single consolidated forecast for the number of accumulated positive COVID-19 tests, each model was set up to generate its output in a common data format for each of the nine federal states of Austria individually. Our forecasts consisted of time series that start with the number of positive tests at the day of the prognosis committee meeting at 11:59 pm and continued in daily time intervals. The length of the time series, the forecasting scope, varied between 8 and 14 days.

Three averaging procedures were considered to generate the joined forecast. These included (1) the point-wise arithmetic mean and two dynamic weighting procedures

wherein the timeseries for each model contributes with a (2) continuous or (3) discrete weighting function with values proportional to the accuracy of its most recent forecasts; see SI Appendix C.

Confidence intervals (CIs) for the harmonized model are derived from the empirical forecast error. Until end of September CIs where derived from the SIR-X model before the method was refined using the empirical forecast error of the harmonized model. Concrete, we retrospectively evaluate the ratio of the consolidated forecast and the actual total number of new cases since the start of the forecast horizon. This ratio follows a log-normal distribution. The upper and lower limits of the CI are derived from the corresponding percentiles of the empirical distribution of this forecast error.

F. Hospital bed usage model

Hospital occupancy is modeled in a stock-flow approach. Inflow (admission to ICU and normal wards) is calculated as a ratio of the time-delayed number of reported or projected new infected. Outflow (discharge) happens after a fixed amount of days for fixed ratios of patients (e.g. 20% of ICU patients are discharged after eight days). Admission rates are calculated separately for sex and age groups (0-39, 40-59, 60-79, and 80 years and above, respectively) and are scaled in order to fit the current occupancy in all federal states. The scaling parameter (one for each federal state) can thus be interpreted as the hospitalisation rate.

Confidence intervals for the occupancy forecast are calculated from increments of the forecast error. A technical description is given in the SI, Appendix D.

Model parameters were initially extracted from literature [29] and subsequently calibrated to actual data to better fit the observed time series. A subsequent analysis based on March–June inpatient data revealed that the calibrated model parameters correspond with observed average length of stay. Refer to the supporting information for a full list of model parameters.

III. RESULTS

A. Forecasting positive test numbers

We show the results for our rolling forecasts compared with the actual case numbers in Figure 1. For the time period from April 4 to September 25 2020, we performed and harmonized weekly forecasts that are visibly as bundles of lines in Figure 1. The first published forecasts coincided closely with the peak of the first epidemic wave in Austria. This can be seen by a gradual flattening of the curve of cumulative case numbers over April. From May until July, the curve showed a linear growth pattern.

While the models showed a clearly discernible divergence for the first prognosis day, the agreement increased over time. The starting points for the early weekly forecasts occasionally lie below the actual cases due to a substantial amount of very late reporting of cases in these early periods of the epidemic. In the weeks thereafter, agreement amongst the three models is typically stronger than the agreement with the data, meaning if one model over- or underestimated the actual trend, so did the other models.

Week 5				
strategy	weights for mean			RMSE (total cases)
	SIR-X	SEIR	SS	
naive	0.33	0.33	0.33	211.00
continuous	0.56	0.18	0.25	198.96
discrete	0.45	0.22	0.33	203.89
Week 10				
strategy	weights for mean			RMSE (total cases)
	SIR-X	SEIR	SS	
naive	0.33	0.33	0.33	85.33
continuous	0.57	0.21	0.22	79.42
discrete	0.45	0.22	0.33	84.78
Week 15				
strategy	weights for mean			RMSE (total cases)
	SIR-X	SEIR	SS	
naive	0.33	0.33	0.33	113.00
continuous	0.37	0.35	0.28	97.14
discrete	0.39	0.28	0.33	96.22
Weeks 2-15				
strategy	RMSE (weeks 2–15, total cases)			
naive	213.14			
continuous	209.38			
discrete	211.47			

TABLE I. Forecast accuracy with different model harmonization strategies. We consider a naive arithmetic mean (strategy "naive") as well as a "continuous" and "discrete" function that weighs the models according to their recent performance. The table displays weights and errors for three selected weeks and a summary of the first 15 forecasts. Performance-weighted model harmonization strategies only marginally improved the accuracy of the forecasts compared to a naive averaging.

We investigated the performance of different averaging procedures that weigh models according to their past performance in terms of their root mean square error (RMSE), see Methods. The results are summarized in Table I. Performance weighting procedures yielded only a marginal improvement over simple averaging in terms of forecast accuracy.

B. Forecasting bed usage

In Figure 2 we show our rolling forecasts for the number of (A) intensive and (B) normal care beds currently in use for COVID-19 patients. While the case numbers and normal ward occupancy peaked in late March, ICU bed usage peaked about two weeks later in April.

The first forecast for normal care occupancy overshot the observed values. With more available data and bet-

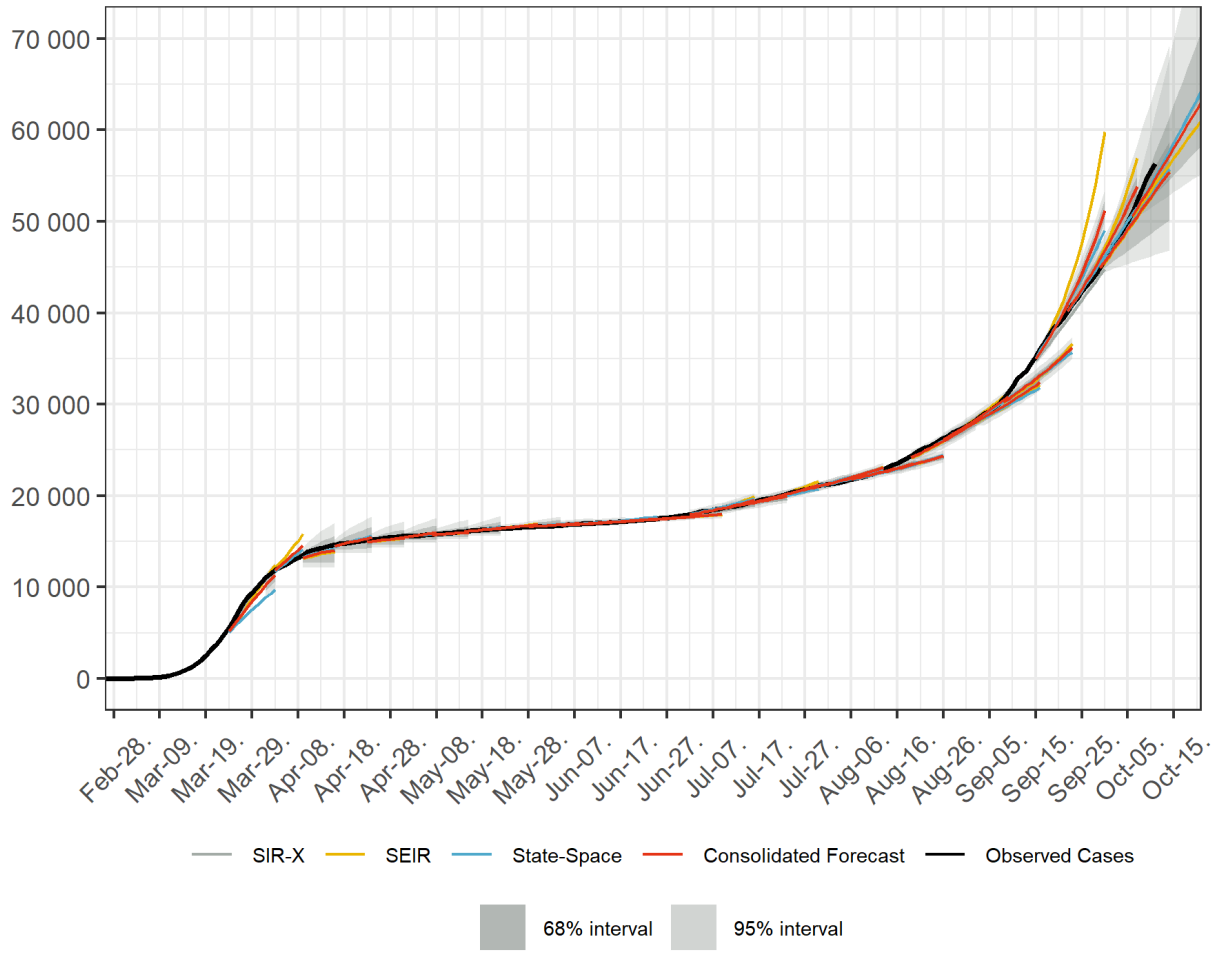

FIG. 1. Rolling combined and consolidated out-of-sample forecasts for the number of confirmed cases in Austria. We show the weekly predictions from the three different models, their arithmetic average with it corresponding CI, and the actual case numbers. In the first two weeks, no CI was given.

ter calibration, the model adequately captured the trends for both normal and intensive care beds. When case numbers started to rise again in late August/September, the model correctly anticipated the corresponding rise in bed usage. In general, ICU occupancy forecasts featured higher accuracy than normal ward occupancy forecasts.

C. What-if scenario results

Using individual models, we considered what-if scenarios to study the impact of NPIs on the infection curve. In particular, we evaluated what would have happened if (i) all NPIs would have been implemented later in time and (ii) the lockdown implemented on March 16, 2020 would not have been taken.

1. *The impact of delayed NPIs*

We evaluated the delayed implementation of all hygiene campaigns and social distancing policies, including closure of stores and schools, cancelling of events, and the restriction of leisure time contact opportunities. Figure 3 displays the simulation results of Agent-Based SEIR Model for a scenario in which these NPIs would have been delayed by 1 to 7 days. The results show that the peak of the confirmed active COVID-19 cases would increase by about 300% if the NPIs would have been implemented 7 days later.

D. No lockdown

Here we consider scenarios in which not all but only measures implemented on March 16 (the lockdown) would have been delayed or not taken at all, see Figure 4. We evaluated scenarios where, all other things be-

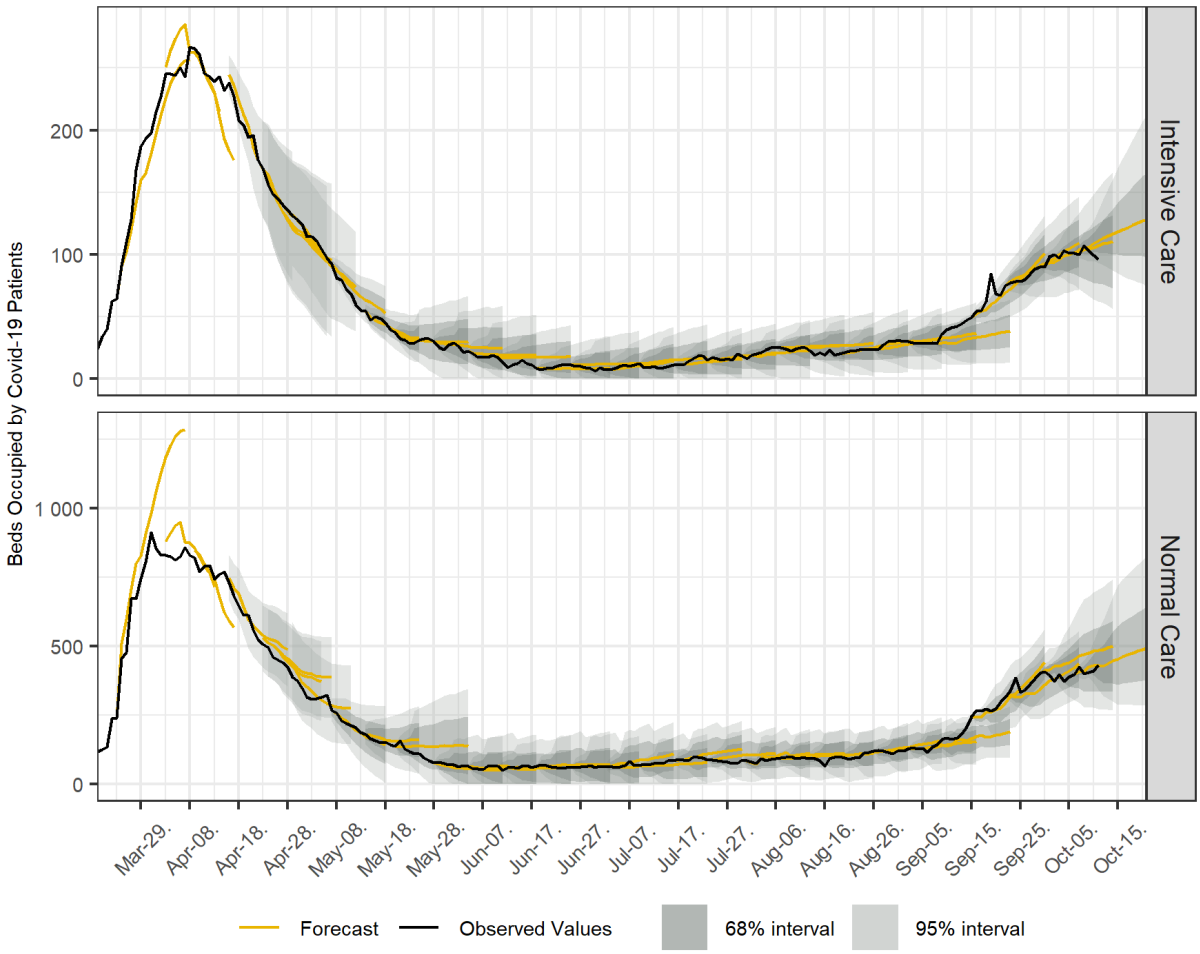

FIG. 2. Rolling combined and consolidated out-of-sample forecasts for the number of (A) intensive and (B) normal care beds currently in use for COVID-19 patients including corresponding CI, and the actual numbers of beds occupied.

ing equal, this lockdown was applied (A) one week later or (B) not at all. By end of April, instead of the approx. 16,000 actual cases, we would have expected 32,000 (95%CI 18,000–51,000) cases would the lockdown have been implemented one week later and 58,000 (95%CI 26,000–114,000) cases in the complete absence of a full lockdown. Assuming a constant case fatality rate, on April 30 these cases would have translated into (A) 1,100 (95%CI 650–1,800) or (B) 2,100 (95%CI 930–4,100) instead of the recorded 549 deaths. With no lockdown taken on March 16, the size of the outbreak in Austria would therefore have been comparable with the dynamics observed in Sweden at the time.

E. Reporting of the forecasts

We developed a standard reporting template used to communicate our forecasts to other stakeholders and decision-makers, see Figure 5. These visual reports showed our forecasts for cases and beds, as well as information on the effective reproduction number and the daily increments in positive tests. The visual reports are complemented by a brief synopsis of the researcher's appraisal of the current situation and particularities of the most recent forecast. We highlight what drives our results and illustrate the nature of the underlying uncertainty, such as the role of tourists returning from risk areas in late August. Furthermore, the researchers are at disposal for any questions that members of the health ministers' office or the regional crisis management units may have.

The first panel in Figure 5 provided an outlook for the expected developments in cases. The trends in these developments, i.e. whether the growth in case numbers accelerates or decelerates, is shown with the timeseries for the daily increments and the effective reproduction number. We communicated the bed forecasts by reporting the prognosis with two CIs, giving the current capacity in an inset. The upper limit of the 95%CI we defined as the "capacity provision", i.e. the upper limit for the number of needed beds at the given level of certainty.

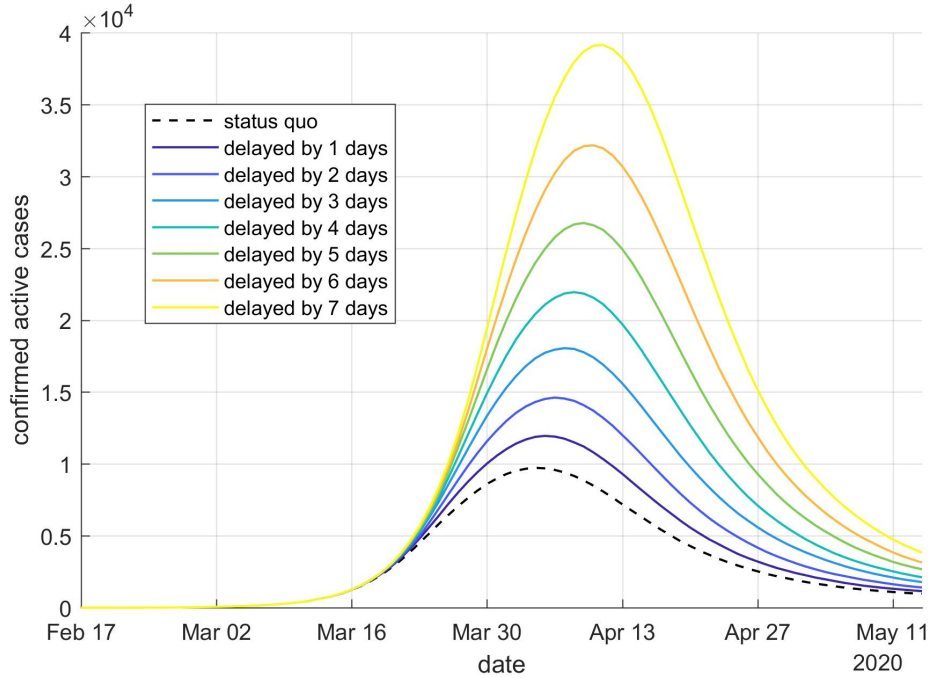

FIG. 3. Agent-Based SEIR Model: Impact of all social distancing NPIs of the Austrian lockdown delayed by up to 7 days on the active confirmed COVID-19 cases.

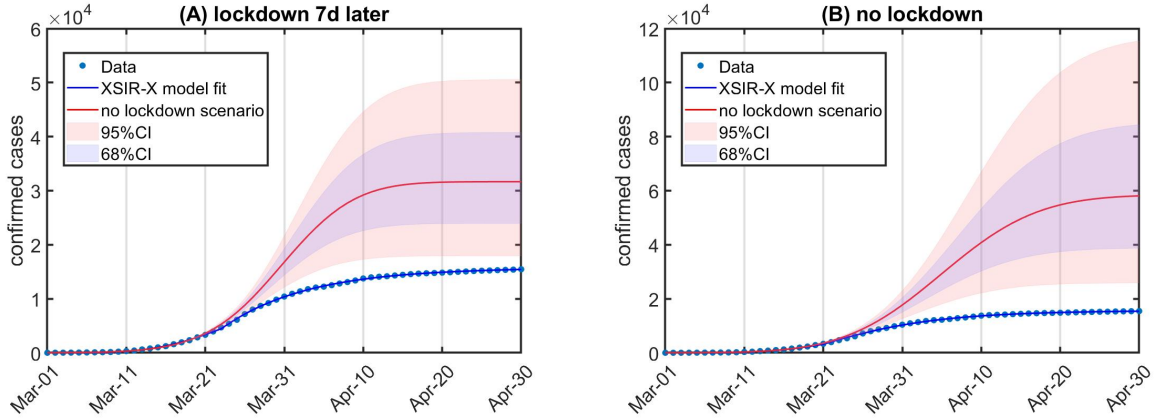

FIG. 4. Extended SIR-X Model: Results for scenarios where only the lockdown from March 16 would have been implemented (A) one week later or (B) not at all. We show the actual data (blue circles) with the model fit (blue lines) and compare it with the scenario simulation (red lines). By end of April, we would have expected around (A) 32,000 or (B) 60,000 instead of 16,000 confirmed cases.

IV. DISCUSSION

Considering the impact of COVID-19 policy measures on economic and social life, any related decision support needs to be done with caution. Our approach considers the high impact of COVID-19 forecasts by (1) focusing on monitoring rather than long term prognosis and (2) consolidation of three different "model opinions" which not only improved the quality of the short term forecast, but also distributed the responsibility of the decision support.

Our forecasts provided sound evidence for the expected number of total and hospitalised cases that include appraisals of uncertainty via forecast intervals. This is in contrast to what has come to be known as "worst case coronavirus science", i.e. the communication of worst case scenarios as baselines in the public pandemic management strategy. For instance, the UK policy change toward adopting more drastic NPIs on March 23 was informed by worst case scenarios created by the Imperial College COVID-19 Response Team that within the current policy regime 250,000 deaths were to be expected.

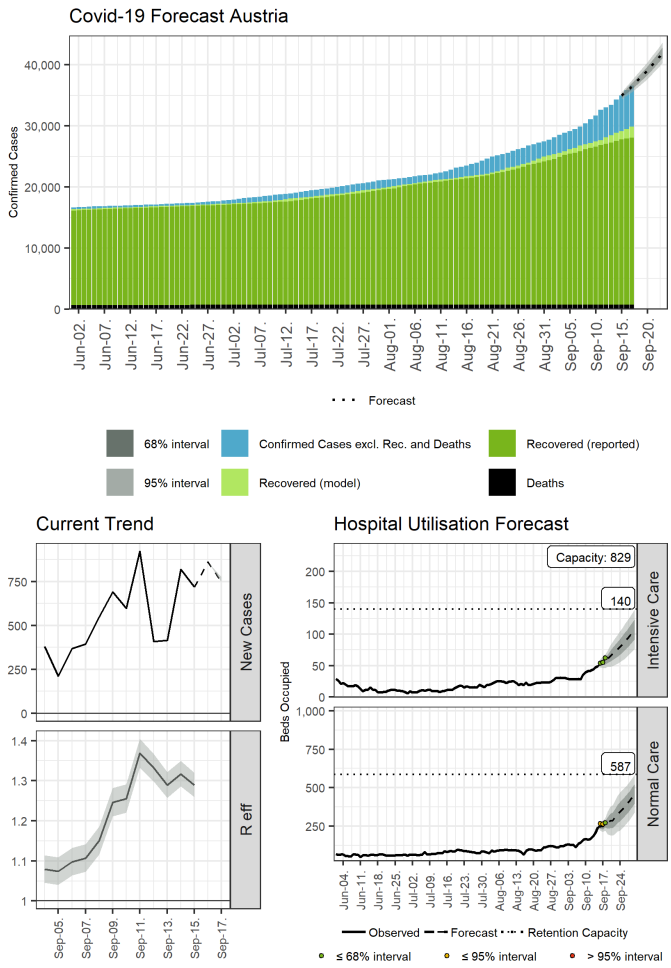

FIG. 5. Example for a reporting template of our out-of-sample forecasts. The visual reports consist of five panels. First, we report the harmonized forecast, its 68% and 95% CI, and the historic case numbers, stratified according to their status of being dead, recovered, or active. Four additional panels show the daily increments in cases, the effective reproduction number, and the forecasts for intensive and normal care beds in use for COVID-19 patients.

In a press conference in April, the Austrian chancellor publicly stated that soon "everyone will know someone who died because of COVID-19", based on an external SIR-model-based worst case scenario that contained a death toll of 100,000 people (1.1% of the Austrian population) [30]. Such scenarios are problematic due to their undefined level of certainty and contributed to a public perception of unreliable epidemiological models. While scenario analysis and numerical experiments in "what if" scenarios have their merits in evaluating effectiveness of certain NPIs, they reduce the trust in government policy once it becomes obvious that the worst case is not going to materialize. If it is the fear of hundreds of thousands of victims in UK or Austria that should compel us to wear masks, why continue wearing them once it is clear that this scenario has been avoided?

Based on our results, we argue that the main benefit of epidemiological models comes from their use as short-term monitoring systems. The models are typically calibrated to the infection dynamics of the last couple of days or weeks and forward project this dynamic based on epidemiological parameters often assumed to be fixed. If a short-term forecast is accurate this means that infection numbers have continued as expected, based on a recent trend. If however, the short term forecast severely over- or underestimates the observed dynamics, one should inquire more closely what might have caused this change. This is in line with our observation that if one of our models considerably underestimated future growth, so did typically the other models that expected a similar trend. Inaccurate short term forecasts therefore signal a change in the epidemiological situation that needs to be explained. Inaccurate predictions can be highly informative from a practical point of view.

Next to its role in monitoring, our approach informed both the strategy to re-open the country and to re-instate measures during the second infection wave. The main insight was that even if a second wave would start the next day, experiences from March told us that we would have several weeks for appropriate reactions (namely re-instating NPIs) before there would be any concern for an overburdening of the healthcare system. For those weeks we could additionally provide estimates for the required number of beds hospitals would need to free for COVID-19 patients, the "capacity provision" for ICU beds. First, it gave hospital managers an estimate for providing a certain number of beds at a given level of risk, thereby freeing up capacities for the treatment of non-COVID patients and minimizing health-related collateral damage. Second, and maybe more importantly for the re-opening, these forecasts clearly revealed that an overcrowding of ICU beds was extremely unlikely from April onward, even in the most pessimistic scenarios (upper limits of the confidence level). Further, the average length of stay for non-COVID-19 patients in ICUs in Austria is less than a week. By postponing non-essential treatments it would therefore be possible to free up additional beds in time if the capacity provision would exceed the number of available beds. Based on such insights, the Austrian government decided to gradually ease the NPIs in intervals of 14 days.

By continually benchmarking the actual case numbers against our trend-following short-term predictions, we could offer an unambiguous signal for whether potential rises in case numbers would be reason for concern or not. The capacity provision served as a weekly checkpoint for the epidemiological situation in this sense: as long as it was not exceeded, there were no significant changes in the infection dynamics and the opening could continue.

Our first ICU forecast that severely underestimated the actual development occurred in September. During the summer, infection numbers increased from around 20 confirmed cases per day to about 200 cases, mostly driven by patients aged below 40y. Consequently, the number

of severe COVID-19 cases remained low and the effective rate to require intensive care dropped to one percent and below. The situation changed qualitatively in September, when not only case numbers started to soar again, but also bed usage increased much more strongly than projected. Our analysis revealed that the driver behind above-forecast ICU occupancy was above-forecast total case numbers, while age-specific ICU rates remained constant. In other words, the bed usage forecasts were inaccurate because of the infection number forecast, but not because the characteristics of the detected cases changed (e.g., more symptomatic or severe cases). This result reassured crisis managers that an uncontrolled spread of the disease, characterised by an increase in undetected cases and, subsequently, an increase in ICU rates, was not happening. What drove this change was a shift of the age distribution of cases from younger patients toward a demographic more representative of the Austrian population. This development was amplified by several non predictable infections in nursing care homes. The fact that our forecast severely underestimated the actual developments, combined with similar signals from other indicators, contributed to a shift in Austria's policy from re-opening to re-instating NPIs.

Our forecast-based decision support comes with limitations. First of all, the weekly prognosis is based on shared data from the Ministry of Health and the Ministry of Internal Affairs which comes with quality and reporting bias limitations. Moreover, even though the consortium has access to the most accurate and up-to-date data about the epidemic in Austria, a lot of information required for valid epidemiological forecasting is available only with considerable delay, or unavailable, since it is and cannot be measured. For instance, the fraction of undetected persons due to asymptomatic disease progression would be one example for such a variable. Further, our forecast is based on simulation models which are generally subject to errors that come from abstraction and simplification of the real system. Through the harmonized handling of three models with entirely different approaches we attempted to reduce such structural uncertainties. Finally, our decision support framework is mostly limited by its political and public visibility. According to our experience, our forecast was of special public and policy interest in periods of rapid movements but also had a confirmatory effect in times of decreasing case numbers or slow growth with respect to taken policy measures.

In conclusion, we argue that worst-case coronavirus science is maybe interesting as an academic exercise, but modellers need to be more cautious and responsible in communicating the strongly speculative nature of their results to politicians and the public. Even if their limitations are adequately discussed in lengthy reports, what often surfaces in the public discourse are tweets and powerpoint slides showing horrific future case numbers and death tolls. Instead, we argue that short-term epidemiological models can be valuable ingredients of a comprehensive monitoring and reporting system to detect epidemiological change points and thereby inform decisions to ease or strengthen governmental responses.

V. ACKNOWLEDGMENTS

We thank Reinhild Strauss, Gabriela El Belazi, Lukas Richter, Daniela Schmid, and Uwe Siebert for helpful and stimulating discussions.

VI. FUNDING

Our work as COVID-19 Forecast Consortium was financially supported by the Federal Ministry for Social Affairs, Health, Care and Consumer Protection. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

VII. AUTHOR CONTRIBUTIONS

MB, CR, and NP developed and operated the Agent-Based SEIR Model. MZ, LR, FB, and HO developed and operated the Epidemiological Clockwork Model. ST and PK developed and operated the Extended SIR-X Model. MB, PK, MZ, LR and FB developed the hospital bed occupancy model. PK wrote the first draft of the article. MB, MZ, LR, FB and CR contributed to the writing. All authors reviewed and edited the manuscript.

[1] C. L. Correa-Martínez, S. Kampmeier, P. Kümpers, V. Schwierzeck, M. Hennies, W. Hafezi, J. Kühn, H. Pavenstädt, S. Ludwig, and A. Mellmann, "A pandemic in times of global tourism: superspreading and exportation of covid-19 cases from a ski area in austria," *Journal of clinical microbiology*, vol. 58, no. 6, 2020.

[2] P. Kreidl, D. Schmid, S. Maritschnik, L. Richter, W. Borena, J.-W. Genger, A. Popa, T. Penz, C. Bock, A. Bergthaler, *et al.*, "Emergence of coronavirus disease 2019 (covid-19) in austria," *Wiener klinische Wochenschrift*, pp. 1–8, 2020.

[3] A. Remuzzi and G. Remuzzi, "Covid-19 and italy: what next?," *The Lancet*, 2020.

[4] A. De Giorgio, "Covid-19 is not just a flu. learn from italy and act now," *Travel Medicine and Infectious Disease*, 2020.

[5] A. Desvars-Larrive, E. Dervic, N. Haug, T. Niederkrotenthaler, J. Chen, A. Di Natale, J. Lasser, D. S. Gliga,

A. Roux, A. Chakraborty, *et al.*, "A structured open dataset of government interventions in response to covid-19," *Scientific Data*, vol. 7, no. 285, 2020.

[6] "Amtliches dashboard covid19." `https://info.gesundheitsministerium.at/`. Accessed: 2020-09-09.

[7] "Covid-19 information page by ages." `https://www.ages.at/en/wissen-aktuell/publikationen/epidemiologische-parameter-des-covid19-ausbruchs-oesterreich-2020/`. Accessed: 2020-08-18.

[8] "Covid-19 information page of the austrian federal ministry of health." `https://www.sozialministerium.at/Informationen-zum-Coronavirus/Neuartiges-Coronavirus-(2019-nCov).html`. Accessed: 2020-08-18.

[9] S. Thurner, R. Hanel, and P. Klimek, *Introduction to the theory of complex systems*. Oxford University Press, 2018.

[10] D. R. L. Sardo, S. Thurner, J. Sorger, G. Duftschmid, G. Endel, and P. Klimek, "Quantification of the resilience of primary care networks by stress testing the health care system," *Proceedings of the National Academy of Sciences*, vol. 116, no. 48, pp. 23930–23935, 2019.

[11] S. Thurner, P. Klimek, and R. Hanel, "A network-based explanation of why most covid-19 infection curves are linear," *Proceedings of the National Academy of Sciences*, 2020.

[12] M. R. Bicher, C. Rippinger, C. Urach, D. Brunmeir, U. Siebert, and N. Popper, "Agent-based simulation for evaluation of contact-tracing policies against the spread of sars-cov-2," *medRxiv*, 2020.

[13] N. Popper, F. Endel, R. Mayer, M. Bicher, and B. Glock, "Planning Future Health: Developing Big Data and System Modelling Pipelines for Health System Research," *SNE Simulation Notes Europe*, vol. 27, pp. 203–208, Dec. 2017.

[14] M. Bicher, C. Urach, and N. Popper, "GEPOC ABM: A Generic Agent-Based Population Model for Austria," in *Proceedings of the 2018 Winter Simulation Conference*, (Gothenburg, Sweden), pp. 2656–2667, IEEE, 2018.

[15] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in recent confirmed covid-19 cases in china," *Science*, vol. 368, no. 6492, pp. 742–746, 2020.

[16] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, and A. Rinaldo, "Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures," *Proceedings of the National Academy of Sciences*, vol. 117, no. 19, pp. 10484–10491, 2020.

[17] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri, "Modelling the covid-19 epidemic and implementation of population-wide interventions in italy," *Nature Medicine*, pp. 1–6, 2020.

[18] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, *et al.*, "Estimating the effects of non-pharmaceutical interventions on covid-19 in europe," *Nature*, vol. 584, no. 7820, pp. 257–261, 2020.

[19] A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Panos, and B. Steinegger, "A mathematical model for the spatiotemporal epidemic spreading of covid19," *MedRxiv*, 2020.

[20] E. Estrada, "Covid-19 and sars-cov-2. modeling the present, looking at the future," *Physics Reports*, 2020.

[21] `https://www.sozialministerium.at/Informationen-zum-Coronavirus/Neuartiges-Coronavirus-(2019-nCov)/COVID-Prognose-Konsortium.html`.

[22] M. Castro, S. Ares, J. A. Cuesta, and S. Manrubia, "The turning point and end of an expanding epidemic cannot be precisely forecast," *Proceedings of the National Academy of Sciences*, 2020.

[23] W. C. Roda, M. B. Varughese, D. Han, and M. Y. Li, "Why is it difficult to accurately predict the covid-19 epidemic?," *Infectious Disease Modelling*, 2020.

[24] N. P. Jewell, J. A. Lewnard, and B. L. Jewell, "Predictive mathematical models of the covid-19 pandemic: Underlying principles and value of projections," *Jama*, vol. 323, no. 19, pp. 1893–1894, 2020.

[25] I. Holmdahl and C. Buckee, "Wrong but useful—what covid-19 epidemiologic models can and cannot tell us," *New England Journal of Medicine*, 2020.

[26] G. Heiler, T. Reisch, J. Hurt, M. Forghani, A. Omani, A. Hanbury, and F. Karimipour, "Country-wide mobility changes observed using mobile phone data during covid-19 pandemic," *arXiv preprint arXiv:2008.10064*, 2020.

[27] M. Bicher, M. Wastian, D. Brunmeir, M. Rößler, and N. Popper, "Review on Monte Carlo Simulation Stopping Rules: How Many Samples Are Really Enough?," in *Proceedings of the 10th EUROSIM Congress on Modelling and Simulation*, (Logrono, Spain), In Print, July 2019.

[28] http://www.dwh.at/en/projects/covid-19/.

[29] `https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html`.

[30] `https://www.oesterreich.gv.at/dam/jcr:a9ba0dbb-fc05-4b6f-a7cb-ecb8b6842364/Executive%20Summary%20Covid19%20v2.pdf`.

[31] "Sir-x model, event horizon." `http://rocs.hu-berlin.de/corona/docs/forecast/model/`. Accessed: 2020-08-18.

SUPPORTING INFORMATION

Appendix A: Details on the Extended SIR-X Model

The SIR-X model [15] was originally devised to model the transition from exponential to sub-exponential growth due to the implementation of non-pharmaceutical interventions (NPIs). With the entry into containment 2.0, i.e. the taking back of NPIs, a couple of modifications are therefore needed to the original model. In particular, the process of "quarantining" susceptibles has to be added. In the following we give a description of how we extended the baseline SIR-X model in order to implement such processes. We also discuss additional extensions to the model, such as introducing age-structured populations and multiple calibration periods. Note that this document only discusses aspects of the model implementation not described or treated different than in the original model description.

1. Compartments for quarantined infected and susceptibles

Here we give the extended SIR-X model to account for scaling back NPIs. The baseline model includes two different types of NPI. First, there are NPIs that act on the susceptible population (social distancing, home office, etc.). Second, there are NPIs that act on the infected population, in particular an accelerated detection of cases (e.g., testing and contact tracing). Clearly, scaling back of the NPIs affects primarily the first type of NPIs, while it might be reasonable to expect that NPIs targeting the infected population might even increase in effectiveness.

The baseline SIR-X model is of the following form,

$$\partial_t S = -\alpha S I - \kappa_0 S \tag{A1}$$
$$\partial_t I = \alpha S I - \beta I - \kappa_0 I - \kappa I \tag{A2}$$
$$\partial_t R = \beta I + \kappa_0 S \tag{A3}$$
$$\partial_t X = (\kappa + \kappa_0) I \quad . \tag{A4}$$

We now introduce two extensions, namely (i) having two compartments of locked down individuals (susceptibles, X^S, and infecteds, X^I) and (ii) introducing a scaling back of NPIs affecting susceptibles encapsulated in the rate $\kappa_1 \geq 0$. The extended SIR-X model is then of the following form,

$$\partial_t S = -\alpha S I - \kappa_0 S + \kappa_1 X^S \tag{A5}$$
$$\partial_t I = \alpha S I - \beta I - \kappa_0 I - \kappa I \tag{A6}$$
$$\partial_t R = \beta I + \kappa_0 S \tag{A7}$$
$$\partial_t X^I = \kappa I \tag{A8}$$
$$\partial_t X^S = \kappa_0 S - \kappa_1 X^S \quad . \tag{A9}$$
$$\tag{A10}$$

There are two notable differences now. First, the compartment X^I is the *cumulative number of confirmed cases*, it will be used to calibrate the model. It is imperative to note that the model was explicitly designed to make statements concerning X^I. A couple of further straight-forward extensions could be introduced to model, for instance, also recovery within the active X^I compartment. Such extensions would further add to the model complexity while providing no value at all for modelling the development of X^I. Secondly, the compartment X^S is now an explicit model representation of locked down or socially distanced susceptibles. The parameter κ_0 gives the inflow to this compartment from the susceptibles (strength of corresponding NPIs), κ_1 gives the outflow (how fast people increase their levels of social contacts back to normal).

2. Age structure

We include an age structure in the model in the ususal way. All compartments (S, I, R, X^I, X^S) become vector-valued, so do the rates κ_0, κ_1, κ, and β. The parameter α becomes a matrix α_{ij} giving the likelihood that a susceptible of age group j will be infected by an infected from age group i. Entries in α have been calibrated using mobile phone data, where we assume that they are proportional to the probability that a call will take place between individuals of age group i and j. The spectral radius of α is chosen in accordance with [15].

3. Calibration

As already mentioned, the model is calibrated via the timeseries of the cumulative number of confirmed cases X^I in each federal state of Austria. The basic calibration procedure follows [15], i.e. we solve the model for solutions to the parameters κ and κ_0, as well as the initial condition $I(t = 0)$ via a trust region reflective algorithm (MatLab's lsqnonlin function). Calibration takes place in different time windows that roughly represent the different phases of the epidemic in Austria. The first phase lasts from $t = 0$ to end of March and encompasses the "first wave". With beginning of April, Austria moved into a containment phase characterized by less than hundred new confirmed cases per day. The second calibration phase ends mid-June where the daily cases started to increase again with most days showing more than hundred new cases. The third calibration phase lasts until the end August, when infection numbers started to increase again, after which the fourth calibration phase commenced.

Appendix B: Details on the Epidemiological Clockwork Model

The epidemiological stage model tracks individuals through the stages "infected", "infectious", "reported", "isolated", and "immunized". It aims to isolate the true infection rate via augmenting reported case numbers with time-constant epidemiological parameters and time-varying assumptions on detection rate, isolation rate, and number of imported cases. This infection rate is then extrapolated using exponential smoothing models in order to forecast future case numbers.

1. Data preparation

In a first step, assumptions on detection and isolation rates are made based on the cluster analysis. For example, if a large share of new reported cases is attributed to a single workplace setting where the whole staff was tested, we assume that contact isolation and detection rate are high at the day these cases were reported, and were lower in days before the mass testing. We furthermore subtract sporadically imported cases as reported in the cluster analysis. This serves to isolate domestically acquired infections for the calculation of the infection rate.

In a second step, case numbers are cleared in a nowcasting procedure, where expected delays and weekend effects are accounted for based on historic values of the relevant federal states. We use exponential smoothing to identify seasonality, error and level of the case numbers for each federal state.

In a third step, assumptions on the risk of importing infectious cases are made, i.e. people who have not acquired the infection domestically but increase the number of infectious.. This figure is parameterised to account for size of federal states and adjusted to reflect travel patterns in risk areas. For example, this value was increased to account for a spike in reported cases that were traced back to travellers from Croatia.

2. Model parameters

Fixed parameters

- Duration infected before infectious: $d_1 = 1$ day
- Duration infectious before reported: $d_2 = 1$ day
- Duration infectious for non-severe cases: $d_3 = 6$ days
  Latency period is therefore 2 days. This value was initially set to 3 days but was reduced owing to evidence that average latency period is shorter. Note that transmission relies on both pre-symptomatic detected cases as well as non-detected cases, who may or may not have mild symptoms.

Variable parameters (default values)

- Background infection risk$_t$ = population$/100,000$
- Detection rate: new reported cases$_t$/new total cases$_t$ $= r_1 = 1/5$
- Effectiveness of contact isolation: new isolated cases$_t$/new reported cases$_t$ $= r_2 = 2/5$

Those given values are standard parameterisations that are adjusted to account for specific circumstances. In the standard case, there would be four additional undetected cases for every reported case, and on average two infected people would be quarantined for every reported case.

3. Mathematical model

The model calculates the number of individuals in each stage according to the following difference equations,

$$\text{new infectious cases}_t = \text{new total cases}_{t+1} - \text{new isolated cases}_t - \text{new immunized cases}_t \tag{B1}$$

$$\text{new immunised cases}_t = \text{new total cases}_{t-6} - \text{new isolated cases}_{t-6} \tag{B2}$$

$$\text{infection rate}_t = (\text{new total cases}_{t+2} - \text{sporadically imported cases}_{t+2})/\text{total infectious cases}_t \tag{B3}$$

$$where \text{total infectious} = \sum_0^t \text{new infectious cases} + \text{background infection risk}_t \quad . \tag{B4}$$

4. Trend extrapolation

In a last step, the isolated true infection rate is forecasted in an exponential smoothing model (R, package ‘forecast‘). Owing to the tendency of the disease to progress linear instead of exponential [11], all trends are damped.

5. Strengths and Limitations

In the Epidemiological Stage Model, isolation of the true infection rate is by a large degree driven by researcher assumptions on the key time-varying variables of imported cases and detection rates. This follows the understanding that reported case numbers should be assessed with available additional information before being further processed by mathematical models. Known sources of error, such as non-detection of cases, travel activity of infected or non-randomised testing will affect, unless these sources of bias are time-constant and the relevant sample size is large. The otherwise simple mechanic of the model facilitates attribution of case numbers to such causes as time of reporting determines day of infection. Of course, researchers may err as well as model when trying to attribute observed patterns to causes, or parameters. The Epidemiologic Clockwork Model will perform better than more data-driven models if known particularities in the time-series of reported cases play a substantial role in the course of transmission. A limitation of the model is that due to the nature of the information used in assessing the time series of reported cases, which is often qualitative in nature, the data clearing process is not transparent. Ongoing improvements of the model will address these shortcomings as the pandemic progresses.

Appendix C: Combining the three models into a consolidated forecast

In summary, three strategies have been evaluated in terms of their forecast error. Let $F_i^j(t)$ denote the forecasts for the total number of COVID-19 cases on day t for model j for runs made in week i,

$$F_i^h(t) =_{j=1,2,3} a_i^j F_i^j(t) \tag{C1}$$

the harmonized forecast, and R_i the corresponding reported number, then the following three strategies have been investigated to calculate the weights $a_i^j, j = 1 \ldots 3$. Note that, here and in the following, j is an index and not an exponent.

- **Naive average**. This strategy describes a static arithmetic average of the forecasts.

$$a_i^j \frac{1}{3} \forall j. \tag{C2}$$

- **Continuously weighted dynamic mean**. This strategy describes a dynamically weighted average. The weights are determined from the forecasting errors of the previous three weeks.

$$a_i^j = \frac{1}{3} \sum_{k=(i-3)}^{i-1} w_k^j, \quad w_k^j = \frac{\frac{1}{\max(|F_k^j - R_k|, 0.5)}}{\sum_{r=1}^{3} \frac{1}{\max(|F_k^r - R_k|, 0.5)}}. \tag{C3}$$

- **Discrete weighted dynamic mean**. In contrast to the continuous weighting, the weights are determined by a step function as follows,

$$a_i^j = \frac{1}{3} \sum_{k=(i-3)}^{i-1} w_k^j, \tag{C4}$$

$$w_k^j = \begin{cases} 1/6, & |F_k^j - R_k| = \max_r(\{|F_k^r - R_k|\}), \\ 1/2, & |F_k^j - R_k| = \min_r(\{|F_k^r - R_k|\}), \\ & 1/3, \quad \text{else.} \end{cases} \tag{C5}$$

Appendix D: Confidence Intervals

Validation of the model is now conducted by evaluating the model's forecast error in each federal state over the last forty days. We compute the empirical distribution of the daily ratios beteen the fitted and observed numbers of cumulative confirmed cases. This gives us a 68% and 95% confidence interval (CI) for the expected accuracy of the model for the prediction for the next day. Following a similar strategy as reported in the online implementation of the baseline model [31], we than recalibrate the model in the last (third) calibration phase by doing as if the upper and lower bounds of these CIs are the actual data points for the next day. The CIs for the forecast are obtained from the forecasts that start from these "virtual" observations.

For the confidence intervals of the ICU and hospital occupancy we decided to apply a different strategy because the fluctuations of the occupancy numbers played a much higher role for the error than the parameter uncertainty.

We found the strategy on increments of the forecasting errors. Assuming that the difference between the real Y_i and simulated $\tilde{Y}_i$ ICU/hospital occupancy $Y_i - \tilde{Y}_i$ at date t_i is displayed as $Y_{i-1} - \tilde{Y}_{i-1} + X_i$ with iid increments $X_i \sim \mathcal{N}(0, \sigma)$ we may estimate the standard deviation σ of this unknown distribution from previous forecasts. With $Y_i, i \in \{1, \ldots, n\}$ available reported occupancy data points and corresponding $\tilde{Y}_i, i \in \{1, \ldots, n\}$ already performed forecasts, we get

$$\sigma \approx \tilde{\sigma} = \sqrt{\frac{1}{n-2} \sum_{i=2}^{n} \left((\tilde{Y}_i - Y_i) - (\tilde{Y}_{i-1} - Y_{i-1}) \right)^2}, \tag{D1}$$

and assume $X_i \sim \mathcal{N}(0, \tilde{\sigma})$. Since,

$$X_i \sim \mathcal{N}(0, \sigma) \Rightarrow \left(\sum_{i=n+1}^{n+1+k} X_i \right) \sim \mathcal{N}(0, \sqrt{k}\sigma),$$

and

$$Y_{n+1+k} - \tilde{Y}_{n+1+k} = \underbrace{(Y_n - \tilde{Y}_n)}_{=0} + \sum_{i=n+1}^{n+1+k} X_i = \sum_{i=n+1}^{n+1+k} X_i$$

we estimate confidence levels for the k-th forecasting day by multiplying the corresponding percentiles of the standard normal distribution by $\sqrt{k}\tilde{\sigma}$. Note, that simulation and real data are synced for the day of the new forecast.